Metacognition in decision-making:

Exploring age-related changes in confidence

Helen Overhoff
ORCID: 0000-0001-5817-2188

November 2022

Submitted in total fulfilment of the requirements
of the jointly awarded degree of Doctor of Philosophy at the

Melbourne School of Psychological Sciences
The University of Melbourne
&
Department of Individual Differences and Psychological Assessment
University of Cologne

Metacognition in decision-making:

Exploring age-related changes in confidence

Inaugural-Dissertation
zur Erlangung des Doktorgrades

der Humanwissenschaftlichen Fakultät
der Universität zu Köln
nach der Promotionsordnung vom 18.12.2018
&
Melbourne School of Psychological Sciences
The University of Melbourne

vorgelegt von
Helen Overhoff
aus Dülmen

November 2022

Diese Dissertation wurde von der Humanwissenschaftlichen Fakultät der Universität zu Köln im April 2023 angenommen.

Abstract

Metacognition is a fundamental human function that supports goal-directed behaviour. By constantly monitoring and evaluating our decisions we are able to detect errors when they occur and adjust the behaviour accordingly. Metacognitive evaluations can be expressed in ratings of decision confidence or error detection reports. Humans are generally capable of forming well-calibrated estimates of their own performance, yet metacognitive abilities have been shown to be specifically affected by healthy ageing. However, the mechanisms underlying this decline remain poorly understood. This thesis aims to investigate the cognitive processes of age-related changes in perceptual metacognitive performance by combining approaches from the fields of error monitoring and decision confidence. For this, we developed a new paradigm for studying the metacognitive evaluation of errors and correct responses that was feasible for adults of all ages. While recording an electroencephalogram (EEG) and response force, a sample of 65 healthy adults from 20 to 76 years made a series of decisions in a modified version of the Flanker task and subsequently indicated how confident they felt about their decision on a four-point scale.

Across two studies, conducted in the same large sample, I addressed three specific research questions: first, how is metacognitive performance affected by healthy ageing? Second, what are factors contributing to the observed decline in metacognitive performance? And third, how does an age-related decline in metacognitive performance affect subsequent behaviour?

The analysis of behavioural data (Study 1a) showed that metacognitive accuracy declined significantly with older age and that this decline could not be explained by the decline in task performance alone. Independent of age, however, participants adjusted their performance according to their metacognitive evaluation of their previous decision and responded more cautiously after reporting low confidence.

The analysis of electrophysiological data (Study 1b) focussed on the modulation of two correlates of error monitoring by confidence and age. The results indicated that the error/correct positivity ($P_{e/c}$), a component discussed as a marker of error detection and decision confidence, scaled with reported confidence in errors but did not show the expected modulation by age. The amplitude of the error/correct negativity ($N_{e/c}$), a marker of early error monitoring processes, also scaled with reported confidence in errors, but in contrast, was less sensitive to variations in confidence with older age.

Finally, Study 2 investigated the effect of age on the relationship between confidence and two response parameters of the initial decision: response time and response force. We replicated a widely reported negative relationship between confidence and response time. Importantly, we showed, for the first time, that confidence was also negatively related to fine-grained changes in peak force, which was intuitively exerted by the participants. Notably, these associations were dependent on the accuracy of the response and changed markedly across age: the relationship between confidence and response time was only found in correct responses and was pronounced with older age, while the relationship between confidence and peak force was only found in errors and only in younger adults.

Overall, these findings jointly provide novel insights deepening our understanding of the observed decline in metacognitive performance with older age. A similar modulation of the $P_{e/c}$ by confidence across the lifespan suggests that the post-decisional process of accumulating evidence about the correctness of a prior decision might generally be intact until old age. Instead, the age-related decline in metacognitive accuracy appears to be related to a multitude of cognitive and neural changes, which might reflect increased noise and hence higher uncertainty in older adults' computation of confidence. Moreover, I discuss how a metacognitive decline could manifest in real life and how recent findings offer a promising view regarding the effect of training on metacognitive performance.

Zusammenfassung

Metakognition bezeichnet eine zentrale Funktion des zielgerichteten Handelns. Die ständige Überwachung und Evaluierung unserer Entscheidungen ermöglicht es uns Fehler zu erkennen, wenn sie auftreten, und unser Verhalten daraufhin anzupassen. Metakognitive Evaluierungen können in Form von Beurteilungen der Entscheidungssicherheit oder von Berichten wahrgenommener Fehler ausgedrückt werden. Im Allgemeinen sind Menschen gut dazu in der Lage, ihr Verhalten richtig einzuschätzen, wobei gezeigt wurde, dass diese spezifische Fähigkeit von nicht-pathologischen Altersprozessen beeinträchtigt wird. Die zugrundeliegenden Mechanismen sind jedoch bisher kaum bekannt. Ziel der vorliegenden Dissertation war daher, durch die Kombination von Ansätzen aus den Forschungsbereichen der Fehlerüberwachung und der Entscheidungssicherheit die kognitiven Prozesse altersbedingter Veränderungen in perzeptiver metakognitiver Leistung zu untersuchen. Dazu wurde ein neues Paradigma zur Erforschung metakognitiver Beurteilungen von Fehlern und richtigen Antworten entwickelt, das für erwachsene Personen jeden Alters durchführbar war. Eine Stichprobe von 65 gesunden Erwachsenen im Alter von 20 bis 76 Jahren traf eine Reihe von Entscheidungen in einer modifizierten Version der Flanker-Aufgabe und gab anschließend auf einer vier-stufigen Skala an wie sicher sie sich ihrer Entscheidung waren. Währenddessen wurde die Kraft der Antworten und mithilfe der Elektroenzephalographie (EEG) die elektrische Gehirnaktivität gemessen.

Über zwei Studien, die in derselben großen Stichprobe durchgeführt wurden, wurden drei spezifische Forschungsfragen untersucht: Erstens, wie wird metakognitive Leistung durch gesundes Altern beeinflusst? Zweitens, welche Faktoren tragen zu der beobachteten Verschlechterung metakognitiver Leistung bei? Und drittens, inwiefern beeinflusst eine altersbedingte Verschlechterung metakognitiver Leistung nachfolgende Verhaltensänderungen?

Die Analyse der Verhaltensdaten (Studie 1a) zeigte, dass die Genauigkeit metakognitiver Beurteilungen mit höherem Alter signifikant abnahm und dass dies nicht allein durch die gleichzeitige Verschlechterung der Leistung in der Entscheidungsaufgabe erklärt werden konnte. Jedoch passten Proband:innen unabhängig von ihrem Alter ihr Verhalten entsprechend der metakognitiven Evaluierung der vorangehenden Entscheidung an und

antworteten vorsichtiger nachdem sie geringe Sicherheit in einer Entscheidung angegeben hatten.

Die Analyse der elektophysiologischen Daten (Studie 1b) konzentrierte sich auf die Modulation von zwei Korrelaten der Fehlerüberwachung durch Entscheidungssicherheit und Alter. Die Ergebnisse zeigten, dass die error/correct Positivity ($P_{e/c}$), eine Komponente, die Fehlererkennung und Entscheidungssicherheit repräsentieren soll, mit der berichteten Sicherheit einen Fehler begangen zu haben variierte, aber nicht die erwartete Veränderung mit dem Alter aufzeigte. Die Amplitude der error/correct Negativity ($N_{e/c}$), einer Komponente der frühen Fehlerverarbeitung, variierte ebenfalls mit der berichteten Sicherheit einen Fehler begangen zu haben, aber war im Gegensatz zur $P_{e/c}$ in höherem Alter weniger sensitiv für Variationen in der Entscheidungssicherheit.

Studie 2 untersuchte schließlich den Effekt von Alter auf den Zusammenhang zwischen Entscheidungssicherheit und zwei Verhaltensparametern der ersten Entscheidung, nämlich der Reaktionszeit und die Reaktionskraft. Wir replizierten einen häufig berichteten negativen Zusammenhang zwischen Sicherheit und Reaktionszeit. Außerdem zeigten wir zum ersten Mal, dass Sicherheit ebenfalls negativ mit kleinsten Änderungen der Reaktionskraft assoziiert war, die natürlicherweise von den Probanden ausgeübt wurde. Entscheidend war, dass diese Zusammenhänge von der Genauigkeit der Antwort abhingen und sich bedeutsam über das Alter hinweg änderten: Der Zusammenhang zwischen Sicherheit und der Reaktionszeit bestand nur in richtigen Antworten und war mit höherem Alter noch stärker ausgeprägt, während der Zusammenhang zwischen Sicherheit und Reaktionskraft nur bei Fehlern und nur bei jüngeren Erwachsenen bestand.

Zusammenfassend tragen diese Befunde zu einem tieferen Verständnis der beobachteten Verschlechterung metakognitiver Leistung mit höherem Alter bei. Eine vergleichbare Modulation der $P_{e/c}$ durch Sicherheit über die Lebensspanne hinweg deutet darauf hin, dass der sich an die Entscheidung anschließende Prozess der Akkumulierung von Hinweisen über die Genauigkeit der vorherigen Entscheidung generell bis in ein hohes Alter intakt zu sein scheint. Stattdessen scheint die altersbedingte Verschlechterung metakognitiver Genauigkeit mit einer Vielzahl kognitiver und neuronaler Veränderungen zusammenzuhängen, die eine erhöhte Verzerrung und größere Unsicherheit älterer Erwachsener in der Einschätzung ihrer Entscheidungssicherheit widerspiegeln. Darüber hinaus wird in dieser Dissertation diskutiert wie sich eine Verschlechterung metakognitiver Fähigkeiten im Alltag äußern könnte und wie jüngste Befunde einen vielversprechenden Blick auf den Effekt von Training auf metakognitive Leistung eröffnen.

Declaration

This is to certify that

(1) this thesis comprises only my original work towards the degree of Doctor of Philosophy, except where otherwise indicated in the preface,

(2) due acknowledgement has been made in the text to all other material used; and

(3) this thesis is fewer than 100 000 words in length, exclusive of tables, figures, references, footnotes, and appendices.

Helen Overhoff

03/11/2022, Helen Overhoff

Preface

I declare that I, Helen Overhoff, have composed this thesis by myself and that this work has not been submitted for any other degree or professional qualification. I confirm that the submitted work is my own, except where jointly authored manuscripts have been included. The contributions of my own and of the co-authors to this work are explicitly indicated below. No third-party editorial assistance was provided in preparation of this thesis.

The data of the studies presented in this thesis was collected by myself with assistance from members of the INM-3 of the Forschungszentrum Jülich, Germany. The empirical Chapter 3 of this thesis is presented in article format: Study 1 comprises an article published by *Neurobiology of Aging* on 8/8/2021 and Study 2 comprises an article submitted to *Frontiers in Aging Neuroscience* on 14/6/2022 that is currently under peer review.

For Study 1, Helen Overhoff, Yiu Hong Ko, Jutta Stahl, Peter H. Weiss, Stefan Bode, and Eva Niessen contributed to the conceptualisation and methodology. Helen Overhoff carried out the data acquisition, guided by Eva Niessen. Helen Overhoff, Yiu Hong Ko, Daniel Feuerriegel, Jutta Stahl, Peter H. Weiss, Stefan Bode, and Eva Niessen contributed to writing the scripts and assisted with analysing and interpreting the data. Helen Overhoff wrote the original draft and all authors reviewed and edited the manuscript. Helen Overhoff is the corresponding author. Jutta Stahl, Peter H. Weiss, Stefan Bode, and Eva Niessen provided supervision. Gereon R. Fink, Peter H. Weiss, Jutta Stahl, and Stefan Bode provided the resources.

For Study 2, Helen Overhoff, Yiu Hong Ko, Jutta Stahl, Peter H. Weiss, Stefan Bode, and Eva Niessen contributed to the conceptualization and methodology. Helen Overhoff wrote the scripts and performed the formal analysis, guided by Eva Niessen. Yiu Hong Ko, Jutta Stahl, Peter H. Weiss, Stefan Bode, and Eva Niessen helped interpret the results. Helen Overhoff wrote the original draft and all authors reviewed and edited the manuscript. Helen Overhoff is the corresponding author. Jutta Stahl, Peter H. Weiss, Stefan Bode, and Eva Niessen provided supervision. Gereon R. Fink, Peter H. Weiss, Jutta Stahl, and Stefan Bode provided the resources.

All co-authors have agreed to the use of these manuscripts in this thesis, and have provided signed copies of the co-author authorisation form. The work in this thesis was supported by

a grant of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation – Project-ID 431549029; SFB 1451) awarded to Gereon R. Fink and Peter H. Weiss-Blankenhorn, an Australian Research Council Discovery Project Grant (DP160103353) awarded to Stefan Bode and a Jülich University of Melbourne Postgraduate Academy (JUMPA) scholarship awarded to Helen Overhoff.

Publications and conference abstracts

Core publications

Overhoff, H., Ko, Y.H., Feuerriegel, D., Fink, G.R., Stahl, J., Weiss, P.H., Bode, S., & Niessen, E. (2021). Neural correlates of metacognition across the adult lifespan. *Neurobiology Aging, 108,* 34–46. https://doi.org/10.1016/j.neurobiolaging.2021.08.001

Overhoff, H., Ko, Y.H., Fink, G.R., Weiss, P.H., Stahl, J., Bode, S., & Niessen, E. (2022). The relationship between response dynamics and the formation of confidence varies across the lifespan. Under review at *Frontiers in Aging Neuroscience*. Preprint available at *bioRxiv:* https://doi.org/10.1101/2022.06.14.493033

Other publications

Ko, Y.H., Feuerriegel, D., Turner, W., **Overhoff, H.**, Niessen, E., Stahl, J., Hester, R., Fink, G.R., Weiss, P.H., & Bode, S. (2022). Divergent effects of absolute evidence magnitude on decision accuracy and confidence in perceptual judgements. *Cognition, 225,* 105125. https://doi.org/10.1016/j.cognition.2022.105125

Conference abstracts

Overhoff, H., Niessen, E., Bode, S., Stahl, J., Ko, Y.H., & Weiss, P.H. (2019). Disentangling Neural Signatures of Error Awareness over the Lifespan, poster presented at the *Students of Brain Research (SOBR) Student Symposium,* Melbourne, Australia.

Overhoff, H., Ko, Y.H., Feuerriegel, D., Fink, G.R., Bode, S., Stahl, J., Weiss, P.H., & Niessen, E. (2020). Effects of ageing on neural correlates of metacognition, talk given at the *MDHS Graduate Research Conference,* Melbourne, Australia (virtual).

Overhoff, H., Ko, Y.H., Feuerriegel, D., Fink, G.R., Bode, S., Stahl, J., Weiss, P.H., & Niessen, E. (2021). Neural correlates of metacognition across the adult lifespan, talk given at the *Decision-making in voluntary action workshop,* London, UK (virtual).

Overhoff, H., Ko, Y.H., Fink, G.R., Stahl, J., Weiss, P.H., Bode, S., & Niessen, E. (2021). Age modulates the effects of response dynamics on the formation of confidence, talk given at the *INM & IBI Retreat,* Jülich, Germany (virtual).

Overhoff, H., Ko, Y.H., Fink, G.R., Stahl, J., Weiss, P.H., Bode, S., & Niessen, E. (2022). Differential relationship between response dynamics and the formation of confidence varies across the lifespan, symposium presentation given at the *64th Tagung experimentell arbeitender Psycholog:innen/Conference of Experimental Psychologists,* Cologne, Germany (virtual).

Overhoff, H., Ko, Y.H., Fink, G.R., Stahl, J., Weiss, P.H., Bode, S., & Niessen, E. (2022). The relationship between response dynamics and the formation of confidence varies across the lifespan, poster presented at the *25th annual meeting of the Association for the Scientific Study of Consciousness (ASSC),* Amsterdam, Netherlands.

Acknowledgements

First of all, I would like to thank Dr. Eva Niessen for her guidance and encouragement, for believing that I would find my way and for her invaluable feedback. You were a fantastic mentor! I would also like to thank my exceptional team of supervisors, Prof. Dr. Jutta Stahl, Prof. Dr. Peter H. Weiss-Blankenhorn, and A/Prof. Stefan Bode for their support and help along the way, and for providing an inspiring research environment throughout my PhD. I am grateful to Prof. Dr. Gereon R. Fink for his advice and for giving me the opportunity to accomplish this work and to be part of the JUMPA program.

I am glad that I could share this entire journey with Yiu Hong Ko, for our fun times at conferences and our mutual support. Furthermore, I could not have done any of this work without two amazing teams which I had the honour to be part of. My special thanks goes to Dr. Claudia Schmidt, Dr. Anne-Sophie Käsbauer, Jane Yook, Dr. Simon Steinkamp, Dr. Max Hommelsen, Dr. Shivakumar Viswanathan, and Dr. Julia Sprenger from the INM(-3) for sharing their knowledge and sharing the office, for (making) train rides (fun), for good food, and for being great role models. I also want to thank Dr. Daniel Feuerriegel, Dr. Milan Andrejevic, Dr. William Turner, Elektra Schubert, Matthew Jiwa, and Philippa Johnson from the DLab for their warm welcome, interesting discussions, fun social times and weekend trips, and to Chiara Van Reyn, Luiza Bonfim Pacheco, and Madeleine Conolly for making me feel home on the other side of the world.

I would like to express my deepest gratitude to my parents, Illa and Peter, and to my grandparents Oma Brigitte and Opa Heinz, and Oma Edith and Opa Willi for always believing in me, in this PhD, and in everything else I do.

Moreover, I am truly grateful to my closest and most reliable companions during the past years and beyond, the DCGs, Angela, Lisa, Madeleine, Valli, and many more for their friendship and for cheering my up, always, but especially during lockdown times in Melbourne. Lastly, I am more than thankful to have Lars by my side, for all of the ways you have helped me and for always making me smile.

Table of Contents

List of Figures

Chapter 1

Chapter 3

Study 1

Study 2

List of Tables

Chapter 3

Study 1

Supplementary Table 28. ANOVA for the prediction of $P_{e/c}$ amplitudes by age and accuracy.

Supplementary Table 29. Post-hoc contrasts for the prediction of $P_{e/c}$ amplitudes by age and accuracy.

Supplementary Table 30. Post-hoc analysis of deviance for the prediction of P_e and P_c amplitudes by age for errors and correct trials.

Supplementary Table 31. ANOVA for the prediction of P_e amplitudes by age and confidence.

Supplementary Table 32. ANOVA for the prediction of P_c amplitudes by age and confidence.

Supplementary Table 33. ANOVA for the prediction of $N_{e/c}$ amplitudes by age and confidence.

Supplementary Table 34. Regression coefficients for the prediction of $N_{e/c}$ amplitudes by age and confidence.

Supplementary Table 35. Post-hoc contrasts for the prediction of $N_{e/c}$ amplitudes by age and confidence.

Supplementary Table 36. Post-hoc analysis of deviance for the prediction of $N_{e/c}$ amplitudes by age for each confidence level.

Supplementary Table 37. ANOVA for the prediction of $P_{e/c}$ amplitudes by age and confidence.

Supplementary Table 38. Post-hoc contrasts for the prediction of $P_{e/c}$ amplitudes by age and confidence.

Supplementary Table 39. ANOVA for the prediction of N2 amplitudes by age and accuracy.

Supplementary Table 40. ANOVA for the prediction of P300 amplitudes by age and accuracy.

Supplementary Table 41. ANOVA for the prediction of raw $N_{e/c}$ amplitudes by age and accuracy.

Supplementary Table 42. Post-hoc contrasts for the prediction of raw $N_{e/c}$ amplitudes by age and accuracy.

Supplementary Table 43. Post-hoc analysis of deviance for the prediction of raw $N_{e/c}$ amplitudes by age for errors and correct responses.

Supplementary Table 44. ANOVA for the prediction of raw $P_{e/c}$ amplitudes by age and accuracy.

Supplementary Table 45. Post-hoc contrasts for the prediction of raw $P_{e/c}$ amplitudes by age and accuracy.

Study 2

Table 1. Regression coefficients for the prediction of confidence in correct responses.

Table 2. Regression coefficients for the prediction of confidence in incorrect responses.

List of Abbreviations

BDI	Beck’s Depression Inventory
cN	centi Newton
CPP	centro-parietal positivity
CSD	current source density
EEG	electroencephalogram
EHI	Edinburgh Handedness Inventory
EOG	electrooculogram
ER	error rate
ERP	event-related potential
fMRI	functional magnetic resonance imaging
(G)LMM	(generalised) linear mixed-effects model
ICA	Independent Component Analysis
M1	primary motor cortex
MMSE	Mini-Mental-State Examination
ms	millisecond
$N_{e/c}$	error/correct negativity
$P_{e/c}$	error/correct positivity
PES	post error slowing
PF	peak force
PFC	prefrontal cortex
PMd	dorsal premotor cortex
RT	response time
SEM	standard error of the mean
TMS	transcranial magnetic stimulation

1. Introduction

In everyday life, we have to make countless decisions in all activities that we engage in. They can be simple (e.g., choosing an ice cream flavour) or complex (e.g., deciding whether it is safe to pass a car on a highway). What they all have in common is that they require to translate some sensory information (e.g., the visual percept of the selection of flavours, or the auditory percept of an approaching car) into an appropriate action of a set of alternatives – perceptual decision-making. While the choice of ice cream flavour will not affect more than the next few minutes of our lives, making the wrong decision on a highway can have serious implications. Therefore, it is essential to continuously monitor and evaluate our behaviour and the associated consequences. Only then are we able to change our decision when we do not feel confident about it, detect errors and quickly adapt the behaviour, and learn to avoid errors in the future. As we age, visual acuity will decrease, processing of information will slow down, and decision-making on the highway will become increasingly demanding. This raises the need for efficient monitoring and adjustment of decision policies, like, for example, realising when driving is not safe anymore.

This ‘meta’ level process of monitoring the own cognitive states (i.e., the ‘object’ level) is referred to as *metacognition*. Metacognition is often described as ‘thinking about thinking’ or ‘cognition about cognition’ and comprises the monitoring and evaluation of ongoing behaviour and its consequences. It is used to guide successful behaviour by adjusting it to the momentary needs and to facilitate learning (Flavell, 1979; Fleming, Dolan, et al., 2012; Heyes et al., 2020).

The overall aim of this thesis is to better understand metacognitive processes in human decision-making, in particular in relation to normal ageing. Acknowledging the increased relevance of metacognition in older adults, I aim to reveal how metacognition, expressed in judgements of confidence or decision accuracy, changes with increasing age. To investigate this, the three studies presented in this thesis will focus on the following questions: *How* is metacognitive performance affected by healthy ageing (Chapter 3, Study 1a)? What are *factors* contributing to the observed decline in metacognitive performance (Chapter 3, Study 1b and 2)? How does an age-related decline in metacognitive performance affect *subsequent behaviour* (Chapter 3, Study 1a)? I note that because all data was acquired together using a single, large sample, the work is presented in relation to these three research questions rather than in the classical format of three separate experiments.

In the following, I will first introduce the research topic of metacognition, how it is operationalised within different domains and conceptualised in computational models of decision confidence, and why it is relevant for successful human behaviour. Next, I will review the literature that is more closely related to the main questions of this thesis, namely effects of ageing on metacognitive performance (Section 1.2), and neural (Section1.3.2) and behavioural (Section 1.3.3) correlates of decision confidence and their potential trajectories with age, situated in the intersection between the fields of error monitoring and decision confidence (Section 1.3.1). Throughout, I will highlight open questions that arise from the existing literature and that will be addressed in my thesis.

1.1. General Background

In this section, I will introduce the concept and significance of metacognition as a field of research. First, I will define key terms and show how metacognition has been operationalised in different cognitive domains. Then, I will briefly outline how computational models have formalised the processes underlying decision confidence and conclude by highlighting the functional significance of accurate metacognitive judgements, which motivates the topic of this thesis.

1.1.1. Quantifying metacognitive performance

Most decisions that we make are based on noisy information, for example, making a choice in a complex environment in daily life (e.g., driving on the highway) or responding to stimuli that are hard to discriminate in an experiment. Moreover, we rarely receive explicit feedback about the accuracy of these decisions. Nevertheless, humans are capable of forming representations of their behaviour that match the objectively observed behaviour remarkably well, yet not perfectly (Fleming & Frith, 2014).

Assessments of metacognitive judgements can be grouped into two categories. They generally require a second-order decision (metacognitive judgement) about a first-order decision (task response). In memory tasks, participants often have to give *prospective* judgements, estimating for example the likelihood that they will remember an item at a later point (judgements of learning) or indicating a feeling that a momentary recall failure will be recognised in the future (feeling of knowing; Hertzog & Curley, 2018). Alternatively, and common in perceptual tasks, metacognition can be assessed *retrospectively*. For example, participants may be asked to make a decision and subsequently report how confident they feel about this decision, or to give a signal when they notice they made a mistake. Different operationalisations of retrospective judgements of confidence and accuracy are illustrated in Figure 1 and will be relevant in the following. For the purpose of this thesis, I will represent confidence on a spectrum from being convinced that a decision was wrong to being convinced that a decision was correct on either ends (Figure 1). I will hence define *low confidence* as the belief of having made an error, while guessing (i.e., not knowing whether the decision was correct or incorrect; Figure 1B) is defined as *low certainty* or *uncertainty* and lies in between low and high confidence. On a discrete rating scale, low certainty can be expressed in different degrees of tendency towards high ('maybe correct') or low ('maybe error') confidence (Figure 1A). Lastly, a metacognitive judgement can also be given as a binary judgement regarding the accuracy of the decision (i.e., whether it was an error or not; Figure 1C). I will refer to a judgement of a decision as being erroneous as low confidence and to a judgement of a decision as correct as high confidence[1].

Furthermore, another retrospective metacognitive assessment that is intended to assign a motivational salience to the assessment is post-decision wagering. Here, participants are asked to place a wager on a prior decision, which should yield higher bets when confidence in a decision is high and lower bets when confidence in a decision is low. It was suggested that this might be a better measure of confidence than directly asking about it and has the advantage that it can be applied in animal research where subjects cannot directly be asked to report their confidence (Fleming & Dolan, 2010; Middlebrooks & Sommer, 2011; for further methods to measure confidence, see Mamassian, 2020).

---

[1] Note that studies in the field of decision confidence often use confidence rating scales that range from guessing to certainty in being correct and refer to guessing as 'low confidence'. In order to capture the full spectrum of metacognitive evaluations, this thesis uses 'low confidence' to express certainty in having made an error, whereas guessing would equate to medium confidence.

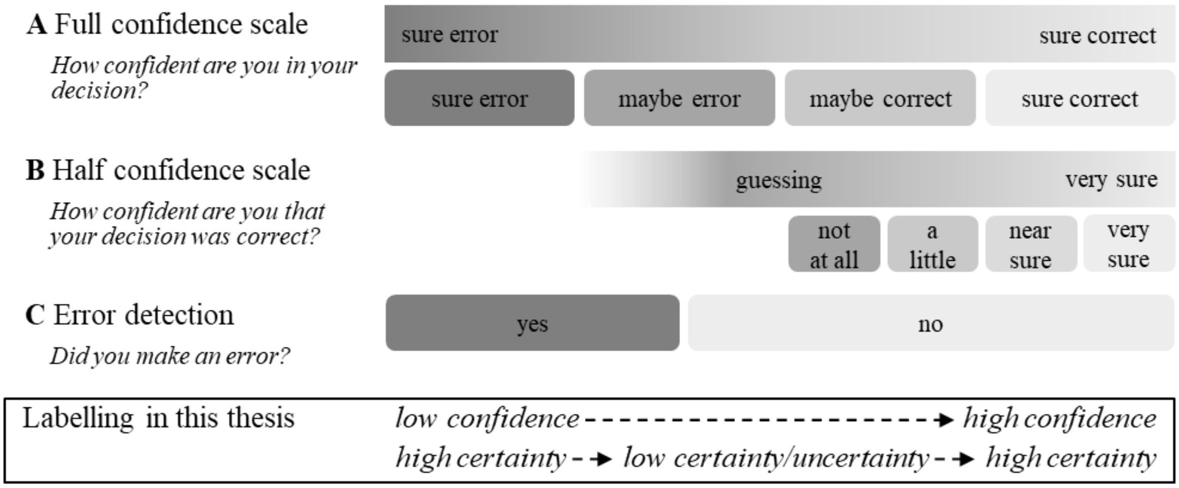


Figure 1. Operationalisations of metacognitive judgements. The metacognitive evaluation of a decision can be assessed in different ways. Participants may be asked to indicate their confidence in a decision on a scale comprising the full spectrum from certainty in having made an error to certainty in being correct (A; e.g., Charles & Yeung, 2019; see also Chapter 3, Study 1 & 2) or the confidence in a correct decision on a scale from guessing to certainty in being correct (B; e.g. Faivre et al., 2018; Rausch et al., 2020) on a continuous or discrete scale, respectively. Alternatively, participants may be asked to make a binary error detection judgement (C; e.g., Masina et al., 2018; or signal an error when by pressing an 'error key'; e.g. Niessen et al., 2017).
For the purpose of this thesis, I will refer to ratings that indicate an error or are at the 'negative' end of a full confidence scale as 'low confidence' or 'high certainty in having made an error', and to ratings that indicate a correct decision or high confidence in being correct as 'high confidence' or 'high certainty in having made a correct decision'. Note, that this is a schematic illustration and definitions of confidence and error detection may vary between studies.

The quality of metacognitive performance is quantified in terms of its *accuracy*, which describes how well a metacognitive report matches the observed outcome of a decision. That is, metacognitive processes are measured at the performance level by assessing the correspondence between a judgement (representing the metacognitive evaluation) and the previously given response (representing the cognitive process that is being monitored). For instance, if a participant correctly decides that the right one of two squares in a discrimination task was brighter than the left one and afterwards reports that the likelihood that this decision was correct is high, this metacognitive evaluation would be *accurate* and for many accurate evaluations, the participant would be assigned a high degree of *metacognitive accuracy*.

When assessed in confidence ratings (i.e., gradually and not as a binary judgement), the ability to discriminate between correct and incorrect decisions is equivalent to *confidence resolution* or *metacognitive sensitivity* of a metacognitive judgement (Fleming & Lau, 2014; Nelson, 1984). To illustrate, metacognitive sensitivity is high when the probability distributions of confidence ratings for correct and incorrect decisions are distant and low when they overlap to a large degree.

Metacognitive *bias*, in contrast, describes a general tendency to give high or low confidence ratings. This means that the mean confidence systematically varies independent of its resolution, that is, confidence ratings might be on average low, but can nevertheless have a high resolution if the ratings well discriminate correct and incorrect decisions (Galvin et al., 2003). As such, people can have a bias towards overall high or low mean confidence (i.e., more frequent use of either the left or the right side of a full confidence scale; Figure 1A), or towards more liberal (i.e., more frequent use of the extreme ends of a full confidence scale; Figure 1A / the right side of a half confidence scale; Figure 1B) or more conservative (i.e., more frequent use of the middle of a full confidence scale; Figure 1A / the left side of a half confidence scale; Figure 1B) confidence ratings (Fleming & Lau, 2014; Steinhauser & Yeung, 2010).

Related to the assessment of metacognition, several measures have been developed to quantify the accuracy of metacognitive judgements in decision-making. When assessed as error detection reports, metacognitive accuracy is represented in the rate of detected errors (Rabbitt, 1966). Correlational measures like Phi (Kornell et al., 2007; Nelson, 1984; for details see Section 4.1.1.1) or the Goodman-Kruskall gamma coefficient G (Kruskal & Goodman, 1954) illustrate the degree to which confidence ratings and decision accuracy overlap and do not require any distributional assumptions. However, these measures are highly susceptible to biases in confidence and to first-order task performance. Therefore, Maniscalco and Lau (2012) developed a model-based approach building on signal detection theory (Green & Swets, 1966), called metacognitive efficiency (meta-d'/d'), that quantifies metacognitive accuracy relative to the performance. Metacognitive efficiency is a relative measure that divides metacognitive sensitivity, the ability to discriminate between correct and incorrect decisions (meta-d'), by the ability to discriminate between stimuli (in a two-alternative forced-choice task; d') and thereby provides a measure of metacognitive performance that is independent of different levels of task performance and metacognitive bias (Fleming & Lau, 2014).

1.1.2. Metacognition across cognitive domains

The accuracy of metacognitive judgements has been investigated across different cognitive domains, of which memory and perception are the most common ones. Research in memory metacognition has a long history (see Hertzog & Curley, 2018 for an overview). It investigates the subjective representation of memory performance, either as a prediction of the ease of later recall or as confidence in accurately remembering something. These types of metacognitive computation are relevant, for example, in the context of learning when you may ask yourself: *Have I revised sufficiently to pass the exam? Is there a topic I just cannot remember and should repeat again?* Despite occasional comparisons to memory metacognition when discussing effects of ageing on metacognitive performance, this thesis will focus on metacognition in perceptual decision-making (in the following "perceptual metacognition").

Within the perceptual domain, metacognition has predominantly been investigated using two different types of judgements: decision confidence and error detection. In the last two decades, the majority of studies using confidence ratings focussed on signal detection tasks (Rahnev et al., 2020). This research concentrates on simple decisions where difficulty is induced by manipulating the quality of evidence. Frequently used paradigms in this field ask participants to discriminate the direction of moving objects (e.g., random dot motion task; Kiani et al., 2014; Resulaj et al., 2009) or compare the evidence between two sets of stimuli (e.g., two patches with different numbers of static dots; Fleming et al., 2016; luminance judgement tasks with flickering stimuli; Y. H. Ko et al., 2022; Turner, Feuerriegel, et al., 2021). By contrast, studies of error monitoring typically use perceptual tasks where the stimuli are easily discriminable, but the quantity of available evidence is manipulated by restricting the time to respond or adding complexity. Prevalent paradigms include tasks that induce conflict (e.g., by presenting distractors that are associated with a different response than the correct one in the Eriksen Flanker task; Eriksen & Eriksen, 1974, by creating mismatch between the spatial locations of a stimulus and the required response in the Simon task; Vissers et al., 2018, by requiring participants to respond to one stimulus property while ignoring another one in the Stroop task; Hajcak & Simons, 2002), require response inhibition (e.g., by inhibiting a response to rare “Nogo” signals in a Go/Nogo task; Niessen et al., 2017, by inhibiting a reflexive saccade in the antisaccade task; Hallett, 1978), or induce difficulty by increasing the response set (e.g., modifications of the flanker task with four or more response options; Maier et al., 2010). To foreshadow, the studies presented in this thesis employed a paradigm that combines the two fields of perceptual metacognition by using a classical error monitoring task and a confidence judgement.

All the aforementioned paradigms tap into the concept of *online* metacognition, which is studied in laboratory experiments and has to be distinguished from *offline* metacognition. Offline metacognition measures cognitive functioning in daily life and can be assessed by questionnaires that are completed as a self-report or by informants of the participant and query various aspects of monitoring and control of behaviour (e.g., beliefs about the own thinking, awareness of learning strategies; Lehmann et al., 2022). While these questionnaires are arguably intended to measure the same latent construct as behavioural measures of online metacognition, growing evidence shows little to no relationship between online and offline metacognition. In fact, self-reports seem to rather capture the general tendency of a participant to give high or low confidence ratings, independent of the performance (Fitzgerald et al., 2017; Lehmann et al., 2022).

1.1.3. Computational models of decision confidence

A number of computational models have been developed to mechanistically describe the processes underlying confidence judgements. A central question of these models is what type of information is integrated into confidence judgements.

One class of models is referred to as *direct access models* and postulates that confidence judgements are based on the same information and cognitive processes that lead to the initial decision (Dotan et al., 2018; Galvin et al., 2003; Kepecs & Mainen, 2012; Kiani & Shadlen, 2009; Shea et al., 2014; Vickers, 1979). In other words, it is assumed that the sensory evidence that informs the first-order decision is likewise used to compute a sense of confidence. These classical theories are derived from the notion that task performance is closely related to metacognitive judgements (Peters, 2022).

One influential framework for direct access models is signal detection theory (Green & Swets, 1966). According to this theory, first-order decisions in two-alternative forced-choice tasks can be categorised into hits, misses, correct rejections, and false alarms. This framework can also be applied to second-order decisions as indicating whether the accuracy of the first-order decision was correctly identified (i.e., rated as correct if it was correct and vice versa; Galvin et al., 2003; Maniscalco & Lau, 2012). While signal detection theory is static, as the probability of being correct defines both the decision and the confidence, the drift diffusion model (Ratcliff & McKoon, 2008) is an example of a framework that models the dynamics of a decision process as the accumulation of evidence over time. In this model, noisy evidence for two response alternatives is accumulated with a certain speed (drift rate) until it reaches a decision threshold. The thresholds for the two response options may be distant or closer together depending on the individual's decision strategy (e.g., they would be more distant if the person accumulates a lot of evidence in favour of one decision until they are absolutely certain that it is correct). Once a decision threshold is crossed, the response is initiated. Additionally, non-decision times like sensory encoding and response execution add to the response time without being part of the decision itself. Applied to second-order decisions, evidence that the initial decision is based on would also be accumulated by the monitoring system regarding the accuracy of the prior decision. Here, the response alternatives could be, for example, high confidence and low confidence or the decision that the initial response was correct or incorrect (Desender, Ridderinkhof, et al., 2021; Pleskac & Busemeyer, 2010).

However, some empirical findings cannot be explained by these models. For example, when comparing trials where sensory evidence remained accessible after the decision and trials where it disappeared at the time of the response, the resolution of the confidence judgements was higher when the participants received ongoing visual information. This was also found when the time between the response and the confidence rating was manipulated and they had more time before being prompted to rate their confidence (Moran et al., 2015). Therefore, it was suggested that the perceptual system continues to process information and feed it to the monitoring system until the time of the confidence judgement. Besides, by investigating movement trajectories, Resulaj and colleagues (2009) showed that participants sometimes changed their minds, that is, after initiating a movement towards one of two possible targets, they changed direction and moved towards the other one instead – without receiving additional information. This is in line with the assumption of post-decisional contributions to metacognition. The authors concluded that the monitoring system may also receive post-decisional information from sensory memory traces that the visual system has not fully processed at the time of the decision. This information, which is still in the processing pipeline, could have affected confidence, which in turn led to changes of mind. These findings have thus been formalised as an extension to a variant of direct access models that allows additional information from post-decisional sensory processing to influence confidence (e.g., Desender, Donner, et al., 2021; Moran et al., 2015; Pleskac & Busemeyer, 2010; Van Den Berg et al., 2016). Crucially, the assumption that confidence continues to emerge after the time of the decision also provides an explanation for effects like changes of mind or error detection, which, by definition, require an initial decision to be made (Charles & Yeung, 2019; Stone et al., 2022). However, the question of when information impacts confidence during initial stimulus processing is not yet resolved (Turner et al., 2022).

A second class of models is based on the counterintuitive finding that confidence can be influenced by information that is not available (or not directly relevant) for the initial decision, for example, information about the motor response indicating the decision (Fleming et al., 2015; Gajdos et al., 2019; Siedlecka et al., 2021; Turner, Angdias, et al., 2021). For example, it was shown that confidence judgements were more accurate after a decision that led to an action compared to decisions that were made by not acting (Siedlecka et al., 2021). Such observations suggest that not only the (ongoing processing of) perceptual information about the stimulus might affect our sense of confidence, but instead various sources of information might be integrated into confidence judgements, which is why these models are referred to as *inferential models* (Shea et al., 2014). It was argued that when sensory information is limited or ambiguous, the monitoring system might also integrate information from new sources that might be informative about the decision accuracy (Fleming & Daw, 2017). Consequently, it was proposed that the metacognitive process of confidence formation might be at least partially distinct from the cognitive process leading to the initial decision (Charles & Yeung, 2019). A recent model of confidence computation provides a theoretical explanation for this by proposing a second-order process of self-evaluation that is separate from the decision process itself and monitors the performance of the first-order decision system (Fleming & Daw, 2017).

While a comparison of these models is beyond the scope of this thesis, a simple distinction helps to understand the theoretical framework in which the presented experiment was conducted, and the latter model in particular provides a rationale for investigating confidence in relation to variables other than stimulus properties, which will be discussed in the context of Study 2.

1.1.4. Functional significance of metacognition and its relation to ageing

When external feedback about the accuracy of a decision is missing, metacognitive evaluations serve as an internal proxy for this information. Effective metacognitive monitoring is crucial for human goal-directed behaviour and should eventually lead to corrections of unfavourable behaviours, learning, and optimisation of decision-making (Fleming, Dolan, et al., 2012; Nelson & Narens, 1990). Only by continuously monitoring and evaluating our actions are we able to identify particularly demanding or non-routine situations, or detect errors. In order to adapt and improve behaviour or mental processes in such cases, cognitive control mechanisms, such as selective attention or inhibition of inappropriate response tendencies, can be applied (Ridderinkhof et al., 2004; Schmidt, 2019).

In the short term, this allows us to quickly adapt the behaviour to the current demands, for example by adopting a more cautious response strategy after the commission of an error. Rabbitt's pioneering work on error processing in the 1960s' reported a relative slowing of responses after errors (post error slowing; PES) that was a sign of an adaptive behaviour to avoid making the same mistake again (Rabbitt, 1966). In a similar vein, it was shown that when participants were uncertain about a decision, they were more likely to choose to see the stimulus again in order to seek additional information and improve the quality of the decisions (Desender et al., 2018). Thus, the metacognitive evaluation of the given situation signalled the need for adapting the behaviour in order to derive a correct decision, suggesting a close link between decision confidence and cognitive control (Desender, Ridderinkhof, et al., 2021). In the longer term, metacognition forms the basis for reasoning and planning by learning from the outcomes of the own behaviour (Metcalfe & Finn, 2008; Yeung &

Summerfield, 2012). Ultimately, metacognition enables the communication of mental states between individuals and hence, plays a crucial role in human social interaction (Heyes et al., 2020).

The importance of metacognition becomes most evident in the context of ageing. We are now faced with a rapid growth of the older population and life expectancy constantly rises. Children born in 2020 have an average life expectancy of 79 (male)/ 83 (female) years in Germany and up to 81 (male)/ 85 (female) years in Australia (*World Bank Open Data*, n.d.). This demonstrates the need to understand cognitive changes that are associated with normal ageing and age-associated neurodegenerative diseases because intact cognition forms the basis of functional independence and communication with others in older age (Murman, 2015). Healthy ageing comes with changes in neural and cognitive processes, related to functional and structural changes in the ageing brain (Harada et al., 2013; Resnick et al., 2003). While crystallized cognition (Lezak et al., 2004), such as knowledge or vocabulary, typically persists or even improves in older age, other aspects of cognition like sensory perception, processing speed, working memory capacity, and executive functions (e.g., decision-making, task-switching) gradually decline. The decline is in general particularly present in functions that depend on prefrontal activity and are related to grey and white matter loss in the prefrontal cortex (PFC) and the medial temporal lobe (Harada et al., 2013; Maillet et al., 2013). Moreover, older age is related to the loss of synapses and functional changes in neuronal networks (Masliah et al., 1993; Rosjat et al., 2020), which can, if they exceed a normal range, predict the development of age-related diseases such as Alzheimer's disease (Bäckman et al., 2005). Of these cognitive changes accompanying ageing, Dully and colleagues (2018) reviewed studies that specifically examined the neural and computational mechanisms underlying perceptual decision-making, which is involved in basically all cognitive tasks. Findings from a small number of studies applying computational models (here: drift diffusion models) of decision-making to ageing suggest that longer non-decision components (i.e., sensory encoding/motor execution) as well as more cautious decision policies that are reflected in increased decision thresholds lead to the prominent slowing of responses in older adults. The prediction of increased non-decision components was validated by neurophysiological studies showing age-related changes in sensory encoding (e.g., in early visual processing) and movement initiation and execution. However, effects of ageing on non-decision components and decision thresholds are task-dependent and their direct influence on the decision-making process remains to be determined. So far, studies examining the effect of ageing on the neural processes underlying the formation of a decision itself are lacking (but see Section 1.3.2).

The sum of the described cognitive changes associated with ageing critically increases the risk for cognitive difficulties and errors in daily life. It has been shown that physical activity might reduce cognitive decline and that the speed and accuracy of decision-making in older adults can be trained (Dully et al., 2018; Murman, 2015). However, this again requires careful monitoring to anticipate potential threats and thus, the ability for effective metacognition. Even though the awareness about increasing cognitive difficulties, for example that the memory slowly fades, can be frightening and concerning in itself, the alternative of being overconfident regarding one's own (cognitive) abilities bears a much larger risk for oneself and others (Castel et al., 2016). While metacognitive judgements are rarely associated with consequences in experiments (but see McGillivray & Castel, 2011, for a positive effect on metacognitive performance of introcuding consequences to confidence judgements in an experiment), in real life, it does make a big difference whether one detects an error or not, or whether one thinks that one will remember something but then does not. Especially in older age, it can be very dangerous if people are not aware of declines in both physical and cognitive abilities. For example, if older adults fail to metacognitively represent growing memory deficits, they may forget to take vital medication or fail to realise their need for specific reminders. Or else, if they fail to detect growing impairments of visual acuity, they may not consider to abandon their driver's licence. Obviously, such lapses can have serious consequences. The next section will outline the current state of research about changes in metacognitive performance accompanying ageing, but more so which open questions remain.

1.2. Metacognition and healthy ageing

In this section, I will review the literature on effects of ageing on metacognitive performance in daily life and psychological experiments. To foreshadow the conclusion, error monitoring studies point to a marked age-related decline in error detection capacity, while studies of age-related changes in metacognition using confidence ratings are still limited.

1.2.1. Metacognitive performance of older adults in daily life

In general, older adults are sceptical about their cognitive abilities. Specifically, besides their belief that cognitive abilities decline with age, they also report having less control about memory and decision-making (Hertzog & Curley, 2018; Rosi et al., 2018). These negative beliefs, however, were shown to rather correlate with psychiatric disorders than with actual problems, and research cautions that complaints or worries about cognitive decline might not always reflect actual awareness of these (Hertzog & Pearman, 2013). In fact, older adults tend to overlook mistakes in daily functioning (Harty et al., 2013; Mecacci & Righi, 2006). When comparing older adults' self-reports about memory or attentional control in everyday situations to their informants' ratings, reports of older adults significantly *over*estimated their abilities, whereas younger adults' reports closely matched those of the informants (Harty et al., 2013). Low awareness of daily functioning was also shown in a prominent study by Ross and colleagues (2012). In this study, participants between 65 and 91 years of age self-rated their driving skills, which were then compared to objective rates of crashes and instances of ignoring traffic regulations. Results showed that older adults highly overestimated their abilities, and self-ratings did not predict the actual driving capacity. Thus, we first need to understand the mechanisms underlying the formation of metacognition in order to understand age-related impairments in these mechanisms, before possible interventions to improve metacognition in older adults can be approached.

1.2.2. Metacognitive performance of older adults in laboratory experiments

In laboratory experiments, effects of healthy ageing on metacognition appear to depend on the cognitive domain being assessed, and research has found functions that are impaired with higher age, while others are spared or even improved compared to younger adults. For example, older adults generally seem to be aware of their decline in memory and show comparable metacognitive performance compared to younger adults for specific aspects of monitoring memory, like judgements about the success of encoding information (Hertzog & Dunlosky, 2011). A recent study compared judgements of learning across different types of memory processes between younger and older adults and showed similar metacognitive efficiency across tasks and age groups despite significantly better performance of the younger adults in the first-order memory tasks (Zakrzewski et al., 2021). Furthermore, older and younger adults showed similar metacognitive abilities in ratings about general knowledge (Dodson et al., 2007) or performance in a problem solving task (Vukman, 2005). Even though research also revealed aspects of memory metacognition that were impaired in older adults (e.g., judgements related to episodic memory or emotionally charged words; Chua et al., 2009; Tauber & Dunlosky, 2012), memory metacognitive abilities seem to remain relatively stable compared to other domains.

In perceptual metacognition, error monitoring studies have consistently revealed a marked decline in older adults' ability to monitor the accuracy of their decisions and signal errors when they occur. Typically, younger and older adults were instructed to make a forced choice in a conflict task and subsequently report errors by giving a signalling response (e.g., pressing one of the response keys for a second time or pressing a designated 'error-signalling' key; e.g., Niessen et al., 2017), by indicating after each trial whether they believed their response was correct or incorrect (e.g., Wessel et al., 2018), or by immediately correcting errors by pressing the actually correct key (e.g., Rabbitt, 1990). Four recent studies, assessing error detection after each trial or through a signalling response using versions of the Flanker, Go/Nogo and antisaccade tasks, found that older adults reported a significantly lower ratio of errors than younger adults (Harty et al., 2017; Niessen et al., 2017; Schreiber et al., 2011; Wessel et al., 2018). Importantly, this was observed even though the overall performance accuracy was similar across age groups or even higher in the older age group (Schreiber et al., 2011). However, surprisingly few studies have hitherto directly assessed error monitoring across the adult lifespan.

Even less is known about the development of metacognitive accuracy in perceptual tasks using confidence ratings. Palmer and colleagues (2014) addressed this issue and compared effects of age on metacognitive efficiency in the two cognitive domains of perception and memory. Sixty participants between 18 and 84 years of age performed a perceptual discrimination task in which they had to identify which of two successive stimulus displays contained a pop-out Gabor patch. Subsequently, they were asked to rate their confidence in the decision on a 6-point scale form uncertainty to high confidence. In a memory task, participants memorised a list of words and later decided on a series of trials which of two displayed words had been contained in the list, again followed by the confidence rating. In order to control for individual differences in task performance, the authors calculated metacognitive efficiency scores for each participant for both perceptual and memory metacognition. Replicating previous findings, metacognitive efficiency in the memory task did not vary as a function of age, while a significant decline was observed for the perceptual task. Neither of these relationships was mediated by age-related changes in executive functions, which could have mediated the effect on metacognition. Accordingly, interindividual, age-related changes in metacognitive performance seem to be domain-specific (i.e., metacognitive accuracy is differentially affected by ageing across cognitive domains; Zakrzewski et al., 2021). One aim of this thesis is to further quantify age-related changes in metacognitive performance in the domain of perceptual decision-making.

1.3. Neural and behavioural correlates of metacognition

In order to shed light on the mechanisms underlying the hypothesised decline in metacognitive performance in older adults, it is important to understand the effects of ageing on the neural and behavioural correlates of metacognition. Up to now, neural correlates of error detection and decision confidence have been studied largely separately. I will instead argue for their shared nature and identify points of convergence while presenting findings of neuroanatomical and electrophysiological correlates of metacognition in each of these fields. Where research findings are available, I will point out how these neural processes vary with age. Finally, I will turn to motor activity as one specific behavioural correlate of metacognition by reviewing a number of recent studies proposing that decision confidence is related to the action of reporting the decision.

1.3.1. Decision confidence and error detection

The concept of metacognition describes the ongoing monitoring and evaluation of behaviour, which can be expressed both in ratings of subjective decision confidence and in error detection reports (Yeung & Summerfield, 2012). Intuitively, asking for the confidence in being correct or the likelihood of having made an error, appear as opposite expressions of the same idea. Interestingly, research in these two arguably related fields has largely been separated. Studies from the field of error monitoring focus on the detection of errors and their impact on behaviour. The employed paradigms, as noted in Section 1.1.1, are conceptually simple tasks, where difficulty is induced by externally imposed conflict or stressors (e.g., Go/Nogo task, Flanker task). Metacognitive assessments ask whether the response was perceived as erroneous and are typically binary judgements, or require a single error indicating signal (Harty et al., 2013; Maier et al., 2008; Wessel et al., 2011). Arising from perceptual decision-making research, the more recent studies of decision confidence, on the other hand, assess metacognition on multi-point or continuous rating scales by asking inversely for the belief in having made a correct choice. These studies mostly draw on intrinsically difficult tasks that require psychophysical discriminations (e.g., orientation discrimination, random dot motion task; Rausch et al., 2020; Resulaj et al., 2009). Here, participants are often uncertain about a response due to stimulus ambiguity (leading to larger variability of confidence ratings in correct responses) but rarely highly certain in having committed an error.

Despite obvious theoretical similarities between the latent constructs of decision confidence and error monitoring, Yeung and Summerfield (2012) pointed to the surprising lack of integration. This may partly be related to fundamental differences in the methodology and theoretical emphasis that impede the translation of results from one area of research to the other. However, several studies recently addressed this issue using rating scales that cover the full spectrum of confidence ranging from certainty in having made an error to certainty in being correct (Boldt & Yeung, 2015; Charles & Yeung, 2019; Pereira et al., 2020; Scheffers & Coles, 2000). For example, Boldt and Yeung (2015) investigated neural markers of decision confidence and error monitoring and provided evidence for a strong link between the two processes. They used a perceptual discrimination task and a confidence assessment on a 6-point scale from ‘certainly wrong’ to ‘certainly correct’. The results showed that an electrophysiological event-related component, previously shown to index error detection, also scaled with graded changes in decision confidence (see Figure 3B). It was therefore claimed that error detection and confidence judgements are part of a continuum of metacognitive evaluation and share underlying neural mechanisms (Boldt & Yeung, 2015). This view has been broadly adopted and was later extended by showing that evidence is similarly integrated into confidence and error judgements over time (Charles & Yeung, 2019). In this thesis, I will adopt the assumption that judgements of decision accuracy and decision confidence are expressions of the same metacognitive process. Therefore, I will use the terms ‘low confidence’ and ‘error detection’ interchangeably.

1.3.2. Age-related changes in the neural basis of metacognition

1.3.2.1. Neuroanatomy of metacognition

Studies using functional or structural magnetic resonance imaging (MRI) have established a central role of the PFC and medial frontal cortex in metacognition in both the non-human (Kepecs et al., 2008; Kiani & Shadlen, 2009) and the human brain (Fleming et al., 2010; Lau & Passingham, 2006). Evidence currently converges on a network of cortical and subcortical regions connected to the posterior medial frontal cortex and adjacent PFC areas to support error monitoring and changes of mind (Bonini et al., 2014; Fleming et al., 2018; Ullsperger et al., 2014), while more anterior prefrontal regions are postulated to be involved in (perceptual) decision confidence (Fleming et al., 2010, 2018). The functional role of the anterior PFC in metacognition was demonstrated by showing that it satisfies key principles of a correlate of decision confidence. Namely, activity in this region was increased when the confidence judgement was freely chosen compared to a condition where the judgement had to be given at a location dictated by the computer. Moreover, PFC activity scaled with reported confidence, and importantly, the stronger the individuals’ relationship between activity and confidence judgements, the higher was their metacognitive accuracy (Fleming, Huijgen, et al., 2012). Finally, Rounis and colleagues (2010) further provided evidence for a causal role of the PFC in metacognition. When delivering bilateral transcranial magnetic stimulation (TMS) to the dorsolateral PFC, the metacognitive accuracy of the participants decreased, while task performance remained unaffected. Taken together, while the specific subregions of the PFC that are activated in a metacognitive task vary slightly and depend on properties of the task (e.g., whether confidence judgement are given pro- or retrospectively; Chua et al., 2009), research has established a central role of the PFC in general in metacognition.

The described brain regions are especially prone to age-related grey and white matter atrophy as well as neurochemical changes (Hämmerer et al., 2014; Salat et al., 1999). Functional neuroimaging studies investigating neural changes in healthy ageing found differences in activation patterns between younger and older adults, characterised by additional recruitment or increased activation of prefrontal areas in older adults (Garavan et al., 2006; Grady, 2012). According to the compensation hypothesis, this over-recruitment of brain activity serves as a compensatory function to maintain a high task performance and thereby working against age-related cognitive deficits (Heuninckx et al., 2008; Reuter-Lorenz et al., 2001). This mechanism is supposed to operate successfully at low levels of task demands only, for example simple perceptual tasks (Chan et al., 2017). In contrast, the dedifferentiation hypothesis is based on the observation that neural structures underlying a specific cognitive function are more broadly distributed in older adults and thus less functionally specialised, leading to the observed increased brain activity when older adults engage in a task (Cabeza, 2001; Rosjat et al., 2020). To date, no consensus has been reached with regards to the leading hypothesis and it is possible that they can also co-exist (Morcom & Henson, 2018). Numerous studies find evidence in favour of each of the theories but also stress that the patterns of brain alterations are not the same for all older adults and that critical events throughout the lifespan may have a larger effect on brain structure and function than chronological age (Reuter-Lorenz & Park, 2014).

1.3.2.2. Electrophysiology of metacognition

A large body of research studied metacognitive processes by recording an electroencephalogram (EEG) of the human scalp. This method is especially useful to identify neurocognitive processes that unfold rapidly in time, and research on neural correlates of metacognition has been quickly growing in the last two decades (Fleming & Frith, 2014; Hämmerer et al., 2014). In the remainder of this section, I will outline and discuss the functional role of two electrophysiological correlates of error monitoring and their age-related trajectories, before turning to another component that has more recently been discussed in relation to decision confidence. While Study 1b of this thesis focussed on the former two components, the latter will be relevant to contextualise the findings reported in Chapter 4 of this thesis.

1.3.2.3. The error/correct negativity

Errors elicit a cascade of cognitive events, of which some are reflected in separable components of the event-related potential (ERP). Research on neural correlates of error monitoring predominantly focussed on two response-locked ERP components: the error negativity (N_e; Falkenstein et al., 1991; Gehring et al., 1993) and the error positivity (P_e; Falkenstein et al., 1991; Nieuwenhuis et al., 2001). The N_e is a negative deflection peaking around 100 ms after the response at fronto-central electrode positions that is larger for first-order errors than correct responses (correct response negativity, N_c, for correct responses; Vidal et al., 2003). The negative component in general ($N_{e/c}$) is assumed to reflect an early post-response evaluation that is sensitive to negative outcomes (Harty et al., 2017; Maier & Steinhauser, 2017). Functionally, different theories postulated that it reflects a mismatch between the executed and the correct response (Coles et al., 2001; Falkenstein et al., 1991), conflict between different response options (Yeung et al., 2004), or prediction errors that trigger a reinforcement learning signal (Holroyd & Coles, 2002). An ongoing debate is whether the $N_{e/c}$ is also a neural correlate of confidence or error detection (Boldt & Yeung, 2015; Rausch et al., 2020). Early studies reported positive correlations between the N_e amplitude and error detection (Falkenstein et al., 1991; Scheffers & Coles, 2000; Yeung et al., 2004), while these findings were subsequently challenged by studies showing no association between the N_e and error detection (Endrass, Schreiber, et al., 2012; Nieuwenhuis et al., 2001). A broad review on the functional role of the N_e concluded that it might serve as one piece of input to different systems that give rise to conscious error awareness (Wessel, 2012).

Many studies investigated effects of ageing on the $N_{e/c}$ amplitude and mostly found reduced amplitudes and/or differences between the amplitude of correct and incorrect trials (Endrass, Schreiber, et al., 2012; Falkenstein et al., 2001; Harty et al., 2017; Larson et al., 2016; Mathewson et al., 2005). Results of a study by Harty and colleagues (2017) are depicted in Figure 2A. In older adults, the amplitude of the $N_{e/c}$ (here: ERN) was significantly smaller for correct responses and errors in the applied Go/Nogo task, and also did not differ between detected and undetected errors. This suggests an inefficient monitoring of response accuracy in older adults (Thurm et al., 2020). However, a few studies did not find age-related differences in $N_{e/c}$ amplitudes (Niessen et al., 2017; Thurm et al., 2013).

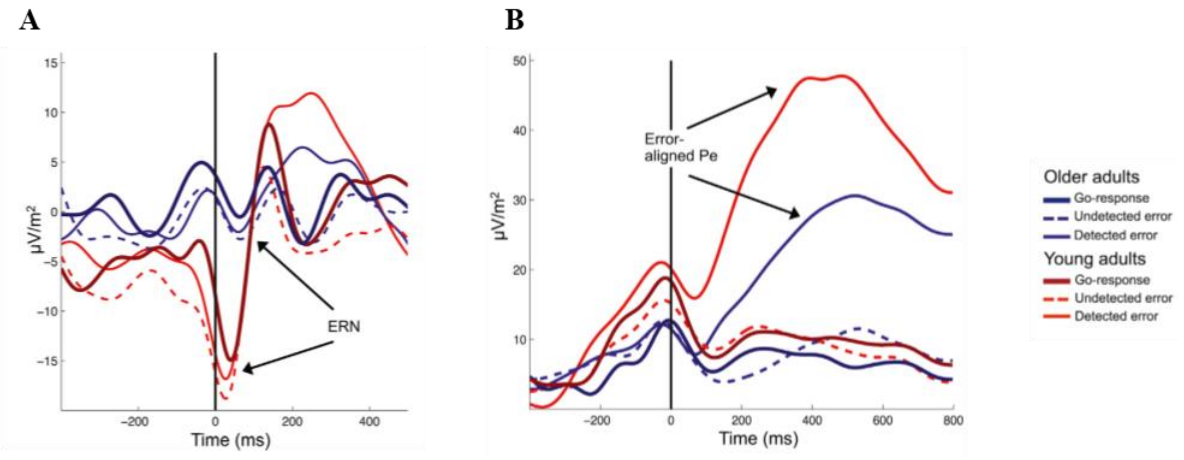


Figure 2. Example of age-related attenuation of $N_{e/c}$ and $P_{e/c}$ amplitudes. A) Response-locked $N_{e/c}$ (here: ERN at FCz) amplitudes did not vary as a function of error detection but were attenuated for all responses for older adults. B) Response-locked $P_{e/c}$ (here: Pe at Pz and P2 for young adults, and POz and PO3 for older adults) amplitudes were significantly larger for detected compared to undetected errors. This difference was reduced for older adults. Adapted from "Parsing the neural signatures of reduced error detection in older age" by S. Harty, P. R. Murphy, I. H. Robertson, and R. G. O'Connell, 2017, *NeuroImage, 161,* p. 49. Copyright 2017 by Elsevier Inc.

1.3.2.4. The error/correct positivity

The second ERP component of interest for studies on error monitoring is the error positivity (P_e; P_c for correct responses). This centro-parietally located, slow positive ERP component with a maximum amplitude around 200-600 ms after the response has been consistently found in detected but not undetected errors or correct responses (Murphy et al., 2015; Nieuwenhuis et al., 2001; Steinhauser & Yeung, 2010; Figure 3A). Consistent with this account, the P_e amplitude has also been shown to correlate with adaptive post-error adjustments (Hajcak et al., 2003; for a review see Danielmeier & Ullsperger, 2011, but see A. Mattes et al., 2022 for a positive correlation with post-error speeding), meaning that the awareness of errors might trigger compensatory mechanisms. Across three experiments, Desender and colleagues (2019) showed that the amplitude of the $P_{e/c}$ predicted the decision policies of the subsequent trial. This is illustrated in Figure 3C as the relationship between $P_{e/c}$ amplitude and subsequent decision bound. It can be seen that higher $P_{e/c}$ amplitudes were related to higher decision bounds, indicating more cautious (i.e., slower) responding in the following decision at a single-trial level.

Effects of ageing on the $P_{e/c}$ have been far less studied compared to the $N_{e/c}$, however, existing error monitoring studies largely coincide on finding a decline in $P_{e/c}$ amplitude for errors in older age (Harty et al., 2017; Larson et al., 2016; Niessen et al., 2017). Harty and colleagues (2017) were explicitly interested in age-related changes in error monitoring and instructed the participants to report via button press whenever they realised an error in the performed Go/Nogo task. In a between-group design with 28 younger (18 - 35 years) and 23 older (65 - 88 years) adults, the older group achieved comparable task performance to the younger group, but reported substantially fewer errors. At the neural level, this was mirrored by alterations in the amplitude and build-up rate of the P_e as can be seen in Figure 2B. These findings extend the well-documented association between P_e amplitude and error detection in young adults to older adults by showing a joint decline and a slowing of the detection process with age. Niessen and colleagues (2017) replicated this finding and additionally showed a gradual decline of error detection and its neural representation in the P_e amplitude across the adult lifespan.

As described in Section 1.3.1, Boldt and Yeung (2015) suggested that the $P_{e/c}$ might mutually index error detection and decision confidence by reflecting metacognitive evaluations on the full continuum from certainty in having made an error to certainty in being correct. In their study, $P_{e/c}$ amplitudes were larger (i.e., they were more positive) when participants were less confident that they were correct when they were actually correct, and also when they were more certain that they had made an error when they had actually committed an error (Figure 3B). A strong negative association between confidence and P_e amplitude was confirmed in a recent study that used a challenging perceptual decision task that resulted in a broad range of reported levels of confidence (Feuerriegel et al., 2022).

Moreover, the P_e was shown to index the accumulation of evidence after a decision, which has been proposed as a computational mechanism underlying the formation of confidence (Desender, Ridderinkhof, et al., 2021; Pleskac & Busemeyer, 2010). In a classical speeded choice task with an error signalling instruction, Murphy and colleagues (2015) studied the morphological characteristics of the P_e. They revealed that the P_e build-up rate and peak latency correlated with the time taken to report an error. Moreover, independent of the response time of the initial decision, the P_e reached a fixed amplitude before the error signalling responses. Together, this qualified the P_e amplitude as a candidate for a neural evidence accumulation signal. This proposed functional role of the P_e can account for findings relating the P_e both to error detection and decision confidence and is summarised in Figure 3 (Desender, Ridderinkhof, et al., 2021).

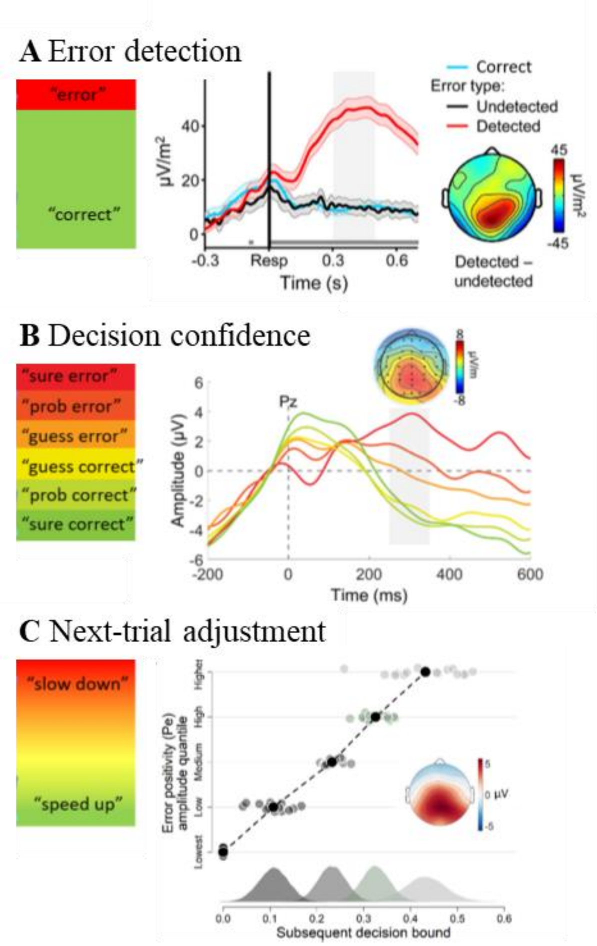


Figure 3. Functional role of the $P_{e/c}$. A) The P_e is larger for detected compared to undetected errors. The individual criterion for reporting an error (i.e., the transition from green to red) can be higher or lower. B) The $P_{e/c}$ increases gradually with lower judgements of decision confidence for errors and correct responses. C) The amplitude of the $P_{e/c}$ predicts the adjustment of response caution in the subsequent trial. Higher amplitudes are related to higher decision bounds in the following trial, resulting in slower and more accurate responses. Adapted from "Understanding neural signals of post-decisional performance monitoring: An integrative review" by K. Desender, K. R. Ridderinkhof, and P. R. Murphy, 2021, *eLife*, *10:e67556*, p. 3. Copyright 2021 by Desender et al.

1.3.2.5. The centro-parietal positivity/P300

Another ERP component that has been studied as a neural marker of perceptual decision-making is the centro-parietal positivity (CPP; O’Connell et al., 2012) or P300 (Twomey et al., 2015). Building on studies on decision formation in non-human primates (Gold et al., 2007; Shadlen & Kiani, 2013), Twomey and colleagues (2015) studied the P300 component of the human scalp ERP, emerging around 300 ms after target stimulus presentation, which has previously been related to a number of cognitive functions and clinical conditions (Polich, 2007), and claimed that it would represent a neural evidence accumulation signal. Based on this finding, the CPP/P300 also qualifies as a potential neural correlate of decision confidence and in fact, it has later been shown to scale with ratings of decision confidence, suggesting that the CPP/P300 does not only track the unfolding of the initial decision but also the formation of second-order metacognitive decisions (Parés-Pujolràs et al., 2020; Rausch et al., 2020).

In other contexts, there has been a large amount of research on lifespan changes in the P300, showing a decrease in amplitude and increase in peak latency with older age (Dully et al., 2018; Polich, 1996). However, these studies often used tasks that differed from the perceptual discrimination tasks used to investigate the CPP in the context of decision confidence and rather investigated the P300 in its role in attentional and memory processes as an index of the reallocation of resources (Polich, 1996, 2007). Thus, the effect of ageing on the CPP/P300 in its relation to confidence has yet to be established.

1.3.3. The influence of action-related information on decision confidence

In Section 1.3.2.1, I have described the brain regions involved in metacognitive evaluations, where several parts of the PFC play a key role (Fleming et al., 2018; Ullsperger et al., 2014). Intriguingly, the lateral PFC has direct connections to the dorsal premotor cortex (PMd), which is known to be involved in the selection of actions (O’Shea et al., 2007). This suggests a potential involvement of the motor system in metacognitive processes and motivated the direct investigation of the relationship between actions and judgements of confidence. In fact, motor activity has been known to play a role in visual perception (Hecht et al., 2001) but most models of perceptual decision confidence did not consider action as a relevant factor contributing to its computation (Anzulewicz et al., 2019). However, the interest in the relationship between action and metacognition has lately grown and is the focus of a number of recent studies on the computation of confidence (Fleming et al., 2015; Palser et al., 2018; Siedlecka et al., 2021; Turner, Angdias, et al., 2021).

Fleming and colleagues (2015) conducted a TMS study on the interaction of confidence and motor-related activity. As the PMd holds action-specific representations and is connected to PFC, it was hypothesised that the disruption of motor-related activity in the PMd might alter confidence. Hence, unilateral single-pulse TMS was applied to the PMd when participants made a response in a visual discrimination task in order to boost or reduce evidence for one of the response options, thereby changing the balance of evidence between the two alternatives. In a control condition, TMS pulses were delivered to the primary motor cortex (M1) to differentiate effects of general motor functions from effects of the cortical representations of action. The results showed that the accuracy of confidence ratings given after the initial decision decreased in the PMd but not in the M1 group, while this perturbation of neural activity did not affect task performance. It was hence proposed that higher-level action representations might contribute to the formation of visual confidence.

Faivre and colleagues (2018) provided analogous electrophysiological evidence for an association between confidence and motor-related information. In a large study investigating metacognition across multiple modalities, they showed significant correlations between confidence judgements and two electrophysiological markers of motor preparation over the sensorimotor cortex (i.e., the lateralised readiness potential; Eimer & Coles, 2003 and alpha power desynchronization; Crone et al., 1998). Stronger motor preparation before the initial response was related to higher confidence in correct decisions. It was suggested that this might reflect attention to parameters of the response execution (e.g., RT), which may inform confidence by providing additional, stimulus-independent information about the decision. As such, sensorimotor electrophysiological activity might contribute to metacognition by moderating the access of the monitoring system to response-related properties of the decision.

A parallel line of research investigated at the behavioural level how motor activity is related to metacognition. In a perceptual discrimination task, Pereira and colleagues (2020) introduced two conditions: in one condition, participants rated their confidence after actively reporting their decisions, while in the other condition, they rated their confidence in decisions that were ‘made’ by a computer hand and only passively observed by the participants. It was shown that confidence ratings predicted decision accuracy significantly better in the active condition and it was argued that metacognitive accuracy may be improved when motor parameters of overt decisions are monitored in addition to the sensory evidence. In the active trials, post-decisional evidence accumulation might be constrained by the decision commitment, leading to less variability in confidence ratings and thus higher metacognitive accuracy in the active compared to the observation condition. Building onto this, Siedlecka and colleagues (2021) followed up on these results by designing an experiment where the decision and the action were not coupled, that is, the condition without action still required the decision to be made by the participant themselves. Again, metacognitive accuracy was higher after decisions that were actively reported compared to those that were reported by not acting, corroborating that information about the motor activity (rather than processes leading to the initial choice) may inform the metacognitive judgement. The authors suggest that the act of reporting a decision may indicate a successful completion of the decisional process and boost confidence.

Together, these results highlight the close link between motor signals and accurate metacognitive evaluation and show that action-related neural activity as well as the action of responding itself influence confidence judgements. Notably, however, the direction of effects differs between experimental designs and focus of analyses. While some studies report an effect of manipulating action-related information on confidence bias (Gajdos et al., 2019; Siedlecka et al., 2020; Turner, Angdias, et al., 2021), others report an effect on metacognitive accuracy (Faivre et al., 2018; Palser et al., 2018), on both bias and accuracy (Siedlecka et al., 2016, 2021), or no effects at all (Filevich et al., 2020). Motor-related information is frequently discussed as sharpening confidence judgements by providing additional information about a decisional process (e.g., its fluency, its completion) besides perceptual evidence. According to the model proposed by Fleming and Daw (2017), the proprioceptive feedback from the motor system, including for example, specifications of movement parameters, might be one source of information that is integrated into the computation of confidence and error detection judgements, which might become relevant when sensory information is limited. If this assumption was true, it is also possible that an impairment of this process contributes to the distorted confidence judgements in ageing. However, the relationship between motor activity and confidence remains to be investigated in the context of ageing.

A recent study further elucidated the proposed association by assessing how a manipulation of the *degree* of action-related information altered decision confidence (Turner, Angdias, et al., 2021). The task required participants to respond with different levels of force, that is, the minimum force to give a response was varied. It was shown that, indeed, confidence scaled gradually with the required force level and also within each level with the exerted force. Importantly, the required force was only revealed after the response had been initiated, which ensured that the exerted effort could not affect the decision itself, but only the associated confidence. This suggests that confidence is not only related to the response as a whole, but that confidence judgements are informed by fine-grained changes in expended physical effort. This thesis aims to first, investigate this relationship in even more detail by quantifying the associations between confidence and two naturally occurring (i.e., not experimentally manipulated) response parameters (i.e., response time and peak force) and second, describe these associations across the adult lifespan.

1.4. The current thesis

The overarching aim of this PhD thesis is to better understand the development of metacognitive performance across the adult lifespan by quantifying age-related changes in the behavioural and electrophysiological correlates of decision confidence. To this end, I will focus on the following research questions: first, *how* is metacognitive performance affected by healthy ageing? Second, what are *factors* contributing to the observed decline in metacognitive performance? And finally, how does an age-related decline in metacognitive performance affect *subsequent behaviour*?

To address the research questions, a novel paradigm was developed that is well suited to test the derived hypotheses. The paradigm is a modified version of the Eriksen Flanker task (Eriksen & Eriksen, 1974), similar to a modification by Maier and colleagues (2008; Maier & Steinhauser, 2017). In order to investigate metacognition in a framework combining theoretical approaches from the decision confidence and error monitoring literature, a conflict task was implemented that is often used in studies of error monitoring. This task was complemented by a metacognitive assessment that allowed to indicate confidence and error detection simultaneously by using a scale that covered the full range of confidence (i.e., a discrete four-point scale comprising the confidence levels ‘surely wrong’, ‘maybe wrong’, ‘maybe correct’, ‘surely correct’). Assessing confidence in correct and incorrect decisions requires a sufficient number of decision errors. This motivated the use of four response options to increase the demand in response selection, and the implementation moderate time pressure. To ensure that the task was feasible and not demotivating, also for older participants who might have perceived the task as more demanding, the following features were implemented based on extensive piloting and prior experience in the lab: we carefully chose the response deadline to be adequate also for participants responding slower on average; we limited the total duration of the experiment to approximately 40 minutes; and we used coloured squares as stimuli to avoid tapping into cognitive functions that were not of primary focus and might be additionally affected by age-related cognitive changes (e.g., visuospatial attention; Curran et al., 2001).

The adoption of this paradigm allowed us to test a large sample of 82 healthy adults until the age of 81 years and to generate a rich dataset that was used for the two studies presented in the empirical Chapter 3 of this thesis. These studies posed the three specific research questions outlined above that are described individually in the following Chapter 2 and will be jointly discussed in Chapter 4. The manuscript of Study 1, comprising research questions 1 to 3, has already been published, and the manuscript of Study 2, which covers research question 2, is currently under peer review.

2. Research objectives

2.1. Study 1a: Effects of ageing on metacognitive performance and its relation to behavioural adaptation

In Study 1a, we investigated changes in metacognition across the adult lifespan, covering research questions 1 and 3. In laboratory tasks, older adults often achieve similar performance levels as younger adults by sacrificing speed for making fewer errors, the so-called speed-accuracy trade-off (Endrass, Schreiber, et al., 2012; Falkenstein et al., 2001; Niessen et al., 2017). The rate of errors that are detected, on the other hand, was consistently found to decrease with age (Harty et al., 2013; Rabbitt, 1990). This has been replicated using different types of perceptual tasks, but metacognitive assessments in these tasks were mostly limited to error signalling or binary error detection reports (Harty et al., 2017; Niessen et al., 2017). A large body of research has studied metacognition in memory and visual perception, however, the computational and neural understanding of metacognitive processes in other types of decision-making tasks as well as studies on individual differences remain at an early stage. Only recently, Palmer and colleagues (2014) examined metacognitive performance in the context of ageing by assessing confidence on a multi-point scale. In healthy adults between 18 and 84 years, they found a gradual decline in metacognitive accuracy despite similar performance in the employed signal detection task. A first aim of Study 1a was to replicate this finding and to examine whether it generalises to a conflict task that is typically used in studies of error detection. In contrast to Palmer and colleagues (2014), our confidence scale allowed to indicate the detection of errors with different degrees of certainty (‘surely wrong’/ ‘maybe wrong’) in order to characterise the effect of healthy ageing on the intersection between error detection and decision confidence. To quantify metacognitive performance, we computed Phi, a correlational measure of metacognitive accuracy that is applicable to four response options and illustrates the degree to which confidence ratings match the observed accuracy (Fleming & Lau, 2014; Kornell et al., 2007; Nelson, 1984).

A second aim of Study 1a was to explore the effects of a potential age-related decline in metacognitive accuracy on subsequent behaviour. Similar to many real-world situations, external feedback about the accuracy of a decision was lacking, so that participants had to use their internal sense of confidence as the best available estimate that adjustments of decision policies could be based on (Desender, Boldt, et al., 2019). Therefore, we examined whether participants adjusted their response tendencies in the subsequent decision according to the level of confidence in a given trial. As the probability of having made an error should be high when confidence is low, an adaptive response strategy would be to respond more cautiously in the following trial, that is, slower and as a result, more accurately. Crucially, we were interested in the development of a potential relationship between confidence and response caution across the lifespan. Behavioural adjustment might either be similarly impaired as metacognitive performance, reflecting more general processing deficits in older adults, or it might be unaffected by ageing, allowing to compensate for potential deficits in metacognitive monitoring. As evidence from error monitoring studies yielded divergent results regarding effects of ageing on adaptive adjustments of behaviour and studies using confidence ratings are still lacking, we remained agnostic regarding the direction of a possible effect of age on the relationship between confidence and response caution (Dutilh et al., 2013; Masina et al., 2018; Niessen et al., 2017).

2.2. Study 1b: Effects of ageing on electrophysiological correlates of metacognition

Related to the research question 2 of uncovering factors contributing to the age-related decline in metacognitive performance, the specific aim of Study 1b was to characterise *neurophysiological* correlates of metacognition and variations thereof related to ageing. For this purpose, two established ERP markers of error monitoring were investigated in their relation to decision confidence. Due to its high temporal resolution, the EEG is well suited to investigate cognitive processing dynamics associated with error processing or metacognition, which consist of a number of cognitive events that occur in a fast sequence (Yeung et al., 2004).

The amplitude of the N_e, a negative component shortly after a response, is typically larger in errors compared to the amplitude of the N_c, its equivalent after correct responses (Falkenstein et al., 1991; Vidal et al., 2003), while the P_e has a larger positive amplitude after errors that are detected compared to errors that are not detected (Falkenstein et al., 1991; Nieuwenhuis et al., 2001). Both components have been discussed in relation to error detection and decision confidence (Boldt & Yeung, 2015; Endrass, Klawohn, et al., 2012; Scheffers & Coles, 2000; Ullsperger et al., 2014), but no consensus has yet been found regarding their specific role in error detection and decision confidence on a single-trial level. Results are mixed with regards to age-related changes of the $N_{e/c}$ (Falkenstein et al., 2001; Larson et al., 2016; Thurm et al., 2013), while the P_e, on the other hand, was consistently attenuated in older adults, in line with lower error detection rates (Harty et al., 2017; Niessen et al., 2017). Somewhat surprisingly, effects of ageing on the $N_{e/c}$ and $P_{e/c}$ have not been investigated in the field of decision confidence to date. Therefore, this question was addressed in Study 1b. We hypothesised that both the $N_{e/c}$ and the $P_{e/c}$ amplitudes would be larger (i.e., more negative and more positive, respectively) for lower compared to higher confidence ratings and expected a specific age-related decline in the $P_{e/c}$ amplitude of lower confidence trials in line with a decline in metacognitive performance.

2.3. Study 2: Effects of ageing on the relationship between confidence and response parameters

The broad aim of Study 2 was to examine one potential source that might contribute to the observed age-related decline in metacognitive performance in Study 1 (research question 2). This objective was addressed by studying two *behavioural* correlates of confidence. Recent studies showed that disrupting action-related information processing or altering movement parameters directly affected confidence ratings and metacognitive accuracy (Fleming et al., 2015; Gajdos et al., 2019), demonstrating an important role of the motor activity of reporting a decision for the formation of confidence. While response time is a well-known correlate of decision confidence in younger adults (Kiani et al., 2014; Rahnev et al., 2020), a recent study showed that confidence is also sensitive to experimentally manipulated changes in the produced force of a response (Turner et al., 2021a). Another recent study provided evidence that metacognition, measured by error detection reports, is also related to naturally occurring levels of force (Stahl et al., 2020). Thus, we systematically examined how these two response parameters, that is, response time (RT) and peak force (PF) were related to confidence judgements, and how these relationships changed with older age. We hypothesised that higher confidence would be related to shorter RT and higher PF. Moreover, given the reported difficulties of older adults to form accurate confidence judgements, we tentatively assumed that the relationship between confidence and the response parameters might be altered with older age, because confidence might be related to the execution of the initial response to a smaller degree than for younger adults.

The underlying idea of Study 2 was that if the action of reporting a decision has a causal effect on metacognition (as suggested by e.g., Kiani et al., 2014; Fleming et al., 2015; Palser et al., 2018), age-related changes in the learnt associations between confidence and response parameters might also contribute to the observed decline in metacognitive accuracy in older adults. However, as our experiment was designed to observe correlational relationships between confidence and the response parameters, and we did not manipulate or reveal the theoretical relevance of the response parameters to the participants, causal inferences were beyond the scope of this study. Nevertheless, the systematic assessment of relevant correlations constitutes the necessary groundwork for future studies to build on and test whether changes in the behavioural correlates of confidence are related to age-related changes in confidence.

3. Publications

3.1. Study 1

3.1.1. Aims

This manuscript presents the equivalent to Study 1a and 1b outlined in the previous section. We explicitly addressed the question of how ageing affects metacognitive performance and related adjustments of subsequent behaviour, and quantified the modulation of electrophysiological correlates of error monitoring by confidence ratings across the lifespan. Metacognitive performance was assessed by computing Phi, a correlational measure of metacognitive accuracy (Kornell et al., 2007; Nelson, 1984), and behavioural adjustments by computing the response caution of a trial in relation to the accuracy and subjective evaluation of the previous trial. This measure provides a combined score of adaptations to accuracy and speed of the subsequent response (Desender, Boldt, et al., 2019). To investigate the neural mechanisms underlying metacognitive processes, we computed single-trial amplitudes of two established markers of error monitoring, the $N_{e/c}$ and the $P_{e/c}$, that have also been discussed in relation to decision confidence. The sample consisted of 65 healthy adults from 20 to 76 years of age who completed our modified version of the Flanker task with four response options, which included a confidence rating on a four-point scale.

This manuscript was published in *Neurobiology of Aging* and is presented in the published format.

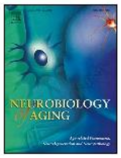
3.1.1. Manuscript

Neurobiology of Aging 108 (2021) 34–46

Contents lists available at ScienceDirect

Neurobiology of Aging

journal homepage: www.elsevier.com/locate/neuaging.org



Neural correlates of metacognition across the adult lifespan

Helen Overhoff[a,b,c,*], Yiu Hong Ko[a,b,c], Daniel Feuerriegel[b], Gereon R. Fink[a,d], Jutta Stahl[c], Peter H. Weiss[a,d], Stefan Bode[b], Eva Niessen[c]

[a] *Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Research Center Jülich, Jülich, Germany*
[b] *Melbourne School of Psychological Sciences, University of Melbourne, Parkville, Victoria, Australia*
[c] *Department of Individual Differences and Psychological Assessment, University of Cologne, Cologne, Germany*
[d] *Department of Neurology, Faculty of Medicine, University Hospital Cologne, University of Cologne, Cologne, Germany*

ARTICLE INFO

Article history:
Received 17 March 2021
Revised 14 July 2021
Accepted 1 August 2021
Available online 8 August 2021

Keywords:
Confidence
Error detection
Aging
Error(-related) negativity (N_e, ERN)
Error positivity (P_e)
Behavioral adaptation

ABSTRACT

Metacognitive accuracy describes the degree of overlap between the subjective perception of one's decision accuracy (i.e. confidence) and objectively observed performance. With older age, the need for accurate metacognitive evaluation increases; however, error detection rates typically decrease. We investigated the effect of ageing on metacognitive accuracy using event-related potentials (ERPs) reflecting error detection and confidence: the error/correct negativity ($N_{e/c}$) and the error/correct positivity ($P_{e/c}$). Sixty-five healthy adults (20 to 76 years) completed a complex Flanker task and provided confidence ratings. We found that metacognitive accuracy declined with age beyond the expected decline in task performance, while the adaptive adjustment of behaviour was well preserved. P_e amplitudes following errors varied by confidence rating, but they did not mirror the reduction in metacognitive accuracy. N_e amplitudes decreased with age for low confidence errors. The results suggest that age-related difficulties in metacognitive evaluation could be related to an impaired integration of decision accuracy and confidence information processing. Ultimately, training the metacognitive evaluation of fundamental decisions in older adults might constitute a promising endeavour.

1. Introduction

We are continuously monitoring and controlling our behaviour in order to achieve goals and avoid errors. The internal evaluation of our behaviour and our decisions, also referred to as *metacognition*, is crucial in everyday life, because it guides our present and future behaviour (Desender et al., 2019b; Rabbitt, 1966). Metacognition comprises both the detection of committed errors and a feeling of confidence that accompanies a decision (Fleming and Frith, 2014; Shekhar and Rahnev, 2020). When we feel less confident about a decision, we might try to adjust it, seek more information, or recruit additional cognitive processes to optimise performance (Desender et al., 2019a, 2019b). As ageing is usually associated with declining cognitive functions and higher rates of decision errors in daily activities, decisions and corresponding motor actions need to be adjusted more often (Hertzog, 2015; Ruitenberg et al., 2014). This might be achieved, for example, by increasing efforts for an efficient metacognitive evaluation of one's behaviour.

In general, metacognitive judgements are highly predictive of actual task performance, yet there is strong evidence that metacognition constitutes a dissociable process from the execution of the initial task (Galvin et al., 2003; Song et al., 2011). The degree to which subjective perceptions and objectively observed performance overlap, that is, the *accuracy* of metacognitive judgements, varies across individuals and task demands (Fleming & Dolan, 2012; Hertzog & Hultsch, 2000; Rahnev et al., 2020). Metacognitive accuracy has been addressed in two separate but arguably related fields of research: studies on error detection, focussing on the recognition of errors, and studies on decision confidence, investigating processes related to beliefs regarding the likelihood of having made a correct choice. In most cases, low confidence implies a higher probability of having committed an error. It has been suggested that error detection and confidence judgements might even share similar underlying computations, whereby error detection arises from low confidence that a correct decision has been made (Boldt and Yeung, 2015; Yeung and Cohen, 2006; Yeung and Summerfield, 2014).

---

[*] Corresponding author at: Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Research Center Jülich, Leo-Brandt-Str. 5, 52425 Juelich, NRW Germany, Phone: +4915161430429

E-mail address: h.overhoff@fz-juelich.de (H. Overhoff).

0197-4580/$ – see front matter © 2021 Elsevier Inc. All rights reserved.
https://doi.org/10.1016/j.neurobiolaging.2021.08.001

1.1. Neural correlates of metacognition

Neural correlates of metacognition have been studied by measuring event-related potentials (ERPs) of the human scalp electroencephalogram (EEG). The error negativity (N_e) is a negative deflection peaking around 100 ms after an overt behavioural response at fronto-central electrodes and typically has larger amplitudes for errors than correct responses (N_c for correct responses; i.e. correct negativity; Falkenstein et al., 1991; Falkenstein et al., 2000; Vidal et al., 2003). The component is classically associated with conflict monitoring, assuming that it tracks conflict between the given response and continuously accumulated post-decision evidence favouring the correct response (Falkenstein et al., 1991; Yeung et al., 2004). Moreover, it has been shown that the N_e amplitude scales with confidence, that is, it decreases from perceived errors to uncertain responses (guesses) to trials where the participant is confident about its correctness (Boldt and Yeung, 2015; Scheffers and Coles, 2000). The more posterior error positivity (P_e; P_c for correct responses, i.e. correct positivity) with a maximum amplitude around 250 ms after a response, is considerably larger for detected compared to undetected errors and has therefore been associated with explicit error awareness (Endrass et al., 2012a; Nieuwenhuis et al., 2001). Notably, the P_e has also been found to increase in amplitude with decreasing confidence in perceptual decisions (Boldt and Yeung, 2015; Rausch et al., 2019).

Concerning the mechanisms underlying these two components, Di Gregorio et al. (2018) designed a sophisticated task to provide evidence that the P_e, but not the N_e, was present when it was evident for participants that an error had been made, but they did not know the correct answer. These findings suggest that the P_e does not require a representation of the correct response to emerge, but instead accumulates post-decisional error evidence from widely distributed neural sources (Di Gregorio et al., 2018; Murphy et al., 2015; Steinhauser and Yeung, 2010; Yeung and Summerfield, 2014). Thus, while both classical components of error processing, N_e and P_e, have been shown to vary with reported confidence, the P_e appears to be more specifically associated with conscious metacognitive processes (Boldt and Yeung, 2015; Nieuwenhuis et al., 2001; Scheffers and Coles, 2000).

1.2. Metacognition and ageing

Metacognitive abilities in older age have been shown to vary across cognitive domains (Fitzgerald et al., 2017; Hertzog & Hultsch, 2000). For instance, while older adults tend to underestimate the prevalence of their decision errors in everyday life, metacognitive judgements of certain memory aspects (e.g., memory encoding) seem to be well preserved (Castel et al., 2016; Harty et al., 2013; Mecacci and Righi, 2006). Previous studies on decision making and metacognition yielded relatively consistent findings of a significant decline in error detection rate with higher age across multiple tasks (Harty et al., 2013; Rabbitt, 1990), even when task performance was comparable (Harty et al., 2017; Niessen et al., 2017; Wessel et al., 2018). In a large sample of healthy adults, Palmer et al. (2014) investigated decision confidence using a measure of metacognitive accuracy that takes task performance into account (Maniscalco & Lau, 2012). The authors found that age was not correlated with metacognitive abilities in a memory task, but that it was negatively correlated with metacognitive abilities in a perceptual discrimination task.

Effects of ageing on the neural correlates of metacognition have primarily been investigated in the field of error detection. Here, both the difference between N_e and N_c (Endrass et al., 2012b; Falkenstein et al., 2001; Schreiber et al., 2011), and the $P_{e/c}$ amplitude (Clawson et al., 2017; Harty et al., 2017; Niessen et al., 2017) were smaller in older adults, while the decrease in P_e, in particular, was linked to a lower error detection rate. Notably, the processing of the stimulus can also affect subsequent response-related processes, and variations with age in two ERPs (namely the N2 and the P300; Groom & Cragg, 2015; Polich, 2007) have been documented (Korsch et al., 2016; Larson et al., 2016; Lucci et al., 2013; Niessen et al., 2017). With the decline in behavioural performance reported above, this suggests an impaired error evidence accumulation process in older age, possibly due to limited cognitive resources (Harty et al., 2017; Niessen et al., 2017). Surprisingly, neither $N_{e/c}$ nor $P_{e/c}$ have been investigated using confidence ratings to assess age-related variations of metacognitive abilities. Some evidence from neuroimaging studies point to age-related structural differences in the neural basis of metacognition (Chua et al., 2009; Hoerold et al., 2013; Sim et al., 2020). However, a conclusive account that explains individual differences in metacognitive accuracy is still missing, for which the use of ERPs with high temporal resolution might be well-suited to provide valuable insights (Dully et al., 2018; Fleming and Dolan, 2012; Yeung and Summerfield, 2014).

1.3. The current study

This study aimed to investigate task performance and metacognition in older adults with a novel perceptual task to determine how generalizable the findings of decreased metacognitive accuracy in older age are (Palmer et al., 2014). For this, we used a colour-flanker task, in which participants had to identify the colour of a target stimulus that was flanked by two squares of the same or a different colour. We assessed decision accuracy, measured confidence using a four-point rating scale, and examined the impact of metacognitive accuracy on adaptations of subsequent behaviour (Desender et al., 2019a; Fleming et al., 2012; Ruitenberg et al., 2014). Furthermore, we investigated whether the amplitudes of $N_{e/c}$ and $P_{e/c}$, which are described as neural correlates of metacognition, track changes in decision confidence across the lifespan.

We hypothesised that metacognitive accuracy in our decision task would decrease with age (Niessen et al., 2017; Palmer et al., 2014). Independent of confidence, we expected an error-specific attenuation of ERP component amplitudes in older adults, which should result in a smaller difference between the neural responses related to errors and correct decisions (Endrass et al., 2012b; Larson et al., 2016). Independent of age, reported confidence was expected to show a positive association with the $N_{e/c}$ and a negative association with the $P_{e/c}$ amplitude (Boldt and Yeung, 2015; Scheffers and Coles, 2000). Based on findings from error detection studies showing an age-related decrease in the P_e amplitude of detected, but not undetected errors (Harty et al., 2017; Niessen et al., 2017), as well as reports linking the P_e to confidence (Boldt and Yeung, 2015), we expected a specific decrease in P_e amplitude for low confidence errors with increasing age.

2. Methods

2.1. Participants

We recruited 82 healthy adults with a broad age range from 20 to 81 years (49.8 ± 1.9 years [all results are indicated as mean ± standard error of the mean; *SEM*]; 35 female, 47 male). We aimed for an approximately uniform distribution of age and thus tested at least 10 participants per decade. Inclusion criteria were right-handedness according to the Edinburgh Handedness Inventory (EDI; Oldfield, 1971), fluency in German, (corrected-to-) normal visual acuity, no colour-blindness and no history of neurological or psychiatric diseases. Any signs of cognitive impairment

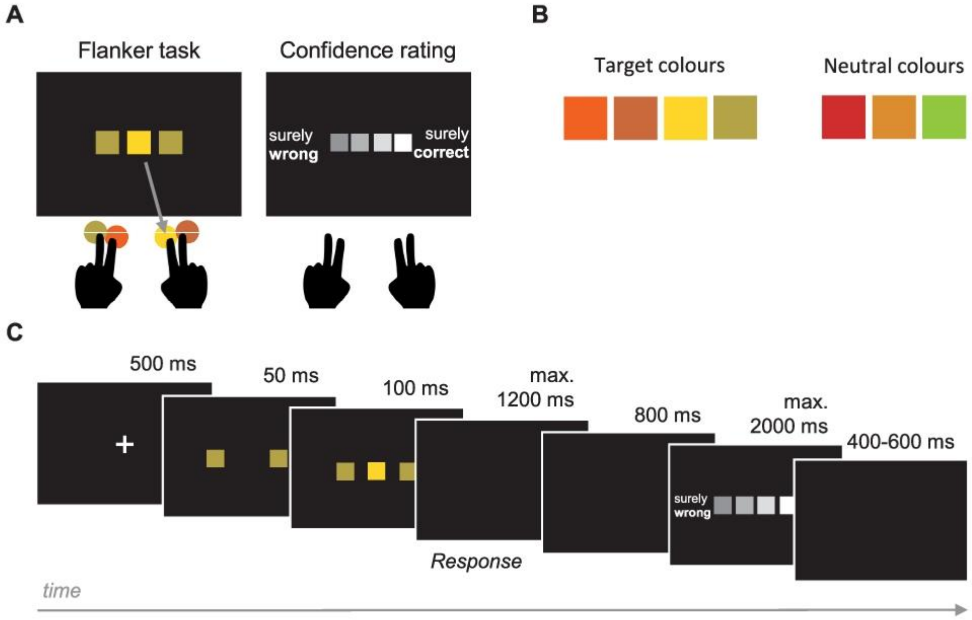


Fig. 1. (A) The left panel shows an example of a trial in the flanker task, where one central target and two flankers were presented, and the participant had to press the finger that was assigned to the respective target colour (illustrated by the grey arrow). The confidence rating (right panel) consisted of four squares, and the ends of the scale were labelled with the German words for 'surely wrong' on the left and 'surely correct' on the right side. The fingers were mapped onto the four squares according to their spatial location. (B) Colours used in the flanker task. Flanker stimuli could consist of target or neutral colours, whereas target stimuli could only consist of one of the four target colours. (C) Sequence of one trial (here, incongruent). Each trial started with a fixation cross, followed by the presentation of the flankers, to which the target was added shortly after. Then, the screen turned black until a response was registered (maximum 1,200 ms), followed by another blank screen. If a response had been given, the rating scale appeared until a rating was given (maximum 2,000 ms). If no response had been given within the designated time window, the German words for 'too slow' were shown instead. The trial ended with another blank screen for a random intertrial interval. For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.

(Mini-Mental-State Examination score lower than 24; MMSE; Folstein et al., 1975) or depression (Beck's Depression Inventory score higher than 17; BDI; Hautzinger, 1991) led to the exclusion of participants (one participant was excluded). Additionally, we excluded four participants who had more than one third of invalid trials (e.g., responses were too slow to fall into the pre-defined response window for analysis, or they showed recording artefacts). Another four participants were excluded because of an error rate (ER) higher than the chance level of 75%. Finally, eight participants were excluded because of combinations of very low accuracy, a high number of invalid trials, the selective use of single response keys, and errors in the colour discrimination task (described below), which suggested a lack of understanding of the task or the use of heuristic response strategies instead of trial-by-trial decisions. After exclusions, the final sample consisted of 65 healthy adults (45.5 ± 2.0 years; 20 to 76 years; 26 female, 39 male).

The study was approved by the ethics committee of the German Psychological Society (DGPs) and conformed to the declaration of Helsinki. All participants gave written informed consent before participating in the experiment.

2.2. Experimental paradigm

The main experimental task consisted of a modified version of the Eriksen flanker task using coloured squares as stimuli and four response options (Eriksen & Eriksen, 1974; Maier & Steinhauser, 2017; Fig. 1A). Participants were asked to respond as fast and accurately as possible to a centrally presented target by pressing a button with one of their index or middle fingers, mapped onto four designated target colours. In each trial, the target consisted of one of these target colours, and the flankers, located right and left to the target, consisted either of the same color as the target (congruent condition), of another target colour (incongruent condition), or of one of three additional neutral colours, which were not mapped to any response (neutral condition [Maier et al., 2008]; Fig. 1B). Both the incongruent and the neutral condition were used to induce conflict as they provided information distinct from the target. We chose this version of the classical flanker paradigm in order to increase task difficulty and thereby maximise the number of errors without tapping into other cognitive processes that might be affected by ageing (e.g., spatial, lexical, or semantic cognition). The colour-finger mapping was fixed over the course of the experiment for each participant and counterbalanced across participants.

Each trial started with a fixation cross for 500 ms. Then, flankers were presented for 50 ms before the target was added to the display for another 100 ms. Showing the (task-irrelevant) flankers before the target was expected to increase the induced conflict (Mattler, 2003). We used a response deadline of 1200 ms because this timing provided a good balance between a desirable number of errors and feasibility for all participants. If no response was registered before this deadline, the trial was terminated and the feedback 'zu langsam' (German for 'too slow') was presented on the screen. If a response was given, a confidence rating scale appeared after a black screen of 800 ms. The delay was introduced to avoid that EEG activity related to the first response overlapped with the confidence assessment. Participants were asked to indicate their confidence in their decision on a four-point rating scale from ‘surely wrong’ to ‘surely correct’ using the same keys as for the initial response. The maximum time for the confidence judgment was 2,000 ms. Trials were separated by a jittered intertrial interval of 400 to 600 ms. The sequence of an experimental trial is depicted in Fig. 1C.

2.3. Procedures

Prior to testing, participants were asked to provide demographic details and complete the handedness questionnaire. Afterwards, they completed a brief colour discrimination task (without EEG) to ensure that all participants were able to correctly discriminate the seven different colours used in the experimental paradigm (see Figure 1B). The discrimination task was followed by the EEG preparation and the main task. The neuropsychological tests were administered after the experiment. In addition, we assessed sustained attention span and processing speed using the d2-test (Brickenkamp, 2002), which have been shown to be positively associated with error processing abilities (Larson et al., 2011).

All stimuli in both tasks were presented on a black screen (LCD monitor, 60 Hz) in an electrically shielded and noise-insulated chamber with dimmed illumination, using Presentation software (Neurobehavioural Systems, version 14.5) for the colour discrimination task and uVariotest software (version 1.978) for the main task. A chin rest ensured a viewing distance of 70 cm to the screen and minimised movements. To record participants’ responses, we used custom-made force-sensitive keys with a sampling rate of 1024 Hz (see Stahl et al., 2020).

The experiment started with a practice block of 18 trials in which participants received feedback about the accuracy of their response, which could be repeated if the participant considered it necessary. After that, two additional blocks with 72 trials without feedback and confidence assessments served as training blocks, allowing the participants to memorise the colour-finger mapping and to get accustomed to the response keys. Afterwards, another practice block introduced the confidence rating to ensure that participants understood and correctly applied the rating scale. The following main experiment consisted of five blocks with 72 trials each. Participants were allowed to take self-timed breaks after each block. The entire session lasted approximately three hours.

2.4. Electroencephalography recording and preprocessing

The EEG was recorded using 61 active electrodes (Acticap, Brain Products) aligned according to the international 10-20 system (Jasper, 1958). The electrodes were online referenced against the posterior Iz electrode close to the inion. Horizontal eye movements were measured using two electrodes at the outer canthi of the eyes (horizontal electrooculogram [EOG]), and another electrode underneath the left eye measured vertical movements (vertical EOG). The EEG signal was recorded continuously at a sampling rate of 500 Hz using a digital BrainAmp DC amplifier (Brain Products). Data were filtered between 0.1 Hz and 70 Hz, and a notch filter of 50 Hz was applied to remove line noise.

EEG data were preprocessed following a standardised pipeline using the MATLAB-based toolboxes EEGLAB and ERPLAB (Delorme and Makeig, 2004; Lopez-Calderon and Luck, 2014). The signal was segmented from -150 to 2,000 ms relative to target stimulus presentation (note that the flankers were presented at -50 ms). Epochs were visually inspected for artefacts and noisy electrodes. Epochs with artefacts were removed and identified noisy channels were interpolated using spherical spline interpolation. To identify and remove eyeblinks, we ran an Independent Component Analysis (ICA) using the infomax algorithm implemented in EEGLAB and afterwards baseline-corrected the epochs using the period of -150 ms to -50 ms to avoid influences of early perceptual processes related to the flanker presentation. Next, data were locked to the response, epoched from -150 ms to 800 ms relative to response onset and baseline-corrected using the 100 ms before the response. The additional analysis of conflict-related stimulus-locked ERPs can be found in the supplementary material S4. Remaining artefacts exceeding ± 150 μV were removed (Niessen et al., 2017), and a current source density (CSD) analysis was conducted using the CSD toolbox (Kayser and Tenke, 2006) allowing for better spatial isolation of ERP components and for obtaining a reference-independent measure (Perrin et al., 1989).

2.5. Behavioural data analysis

Trials with invalid responses (i.e. responses that were too slow) or recording artefacts, as well as responses faster than 200 ms were excluded from further analysis. The error rate (ER) was calculated as the proportion of errors relative to valid responses. Response time (RT) was defined as the time between stimulus onset and the initial crossing of the force threshold (40 cN) by any of the response keys.

To inspect how the confidence scale was used across participants, raw distributions of confidence ratings within all incorrect and correct responses were extracted. We computed Friedman ANOVAs for the proportion of each of each rating level for errors and correct responses with the factor confidence (4 levels). This analysis revealed that only a limited number of trials was available for the two middle confidence rating levels (‘maybe wrong’, ‘maybe correct’), and we therefore collapsed those to create one category for all further analyses representing ‘unsure’ responses, i.e. confidence ratings expressing uncertainty.

For the analysis of metacognitive accuracy, we computed the Phi (Φ) correlation coefficient, which is a simple trial-wise correlation between task accuracy and reported confidence. It describes the extent to which the distributions of confidence ratings for correct and incorrect trials differ, while still depending on primary task performance and individual biases in confidence judgements (Fleming and Lau, 2014; Kornell et al., 2007; Nelson, 1984). Phi was calculated by correlating accuracy, coded as 0 (error) and 1 (correct response), and confidence (that the given response was correct), coded as 1 (‘surely wrong’), 2 (‘unsure’), and 3 (‘surely correct’), for each participant. This provided us with one measure of metacognitive ability per participant that comprises both the accuracy and the confidence rating of each trial (e.g., Phi $= 1$ means that correct trials were successfully identified as such without uncertainty; while a Phi $= 0$ means that all errors were rated as ‘surely correct’, or all correct trials were rated as ‘surely incorrect’).

To assess the impact of accuracy and confidence on trial n on adaptations of behavioural responses, we computed a measure of response caution by multiplying the accuracy and RT on trial $n+1$ (Desender et al., 2019a). Response caution captures the trade-off between speed and accuracy in a decision, with higher values indicating a more cautious response strategy that is characterised by slower, and at the same time, more accurate responses. For this analysis, only pairs of two consecutive valid trials were included. Response caution was computed separately relative to a) initial trial accuracy (error, correct), and b) initial trial confidence (‘surely wrong’, ‘unsure’, ‘surely correct’).

At the group level, age-related effects on the d2-test score, the error rate, and Phi were computed using linear regressions.

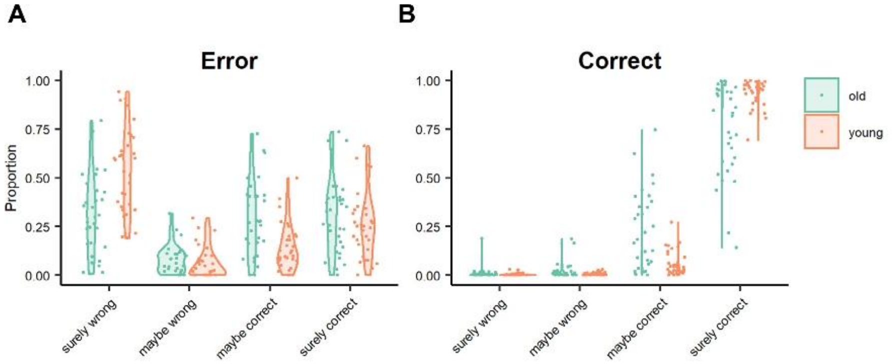


Fig. 2. Distributions of confidence ratings for errors (A) and correct responses (B). Errors were most often rated as ‘surely wrong’, and correct responses as ‘surely correct’. Dots represent the individual proportions of the particular confidence response amongst all errors or correct responses, respectively. A median split by age ($Mdn = 46$) was conducted for illustration purposes. Older adults are shown in green, younger adults in orange. With increasing age, participants used the ‘surely correct/wrong’ ratings less often, and the middle of the confidence scale more often.

To rule out that metacognitive accuracy was confounded by age-related impairments in task performance or attention and processing speed, we performed additional multiple linear regressions to predict Phi by age, adding the factors of error rate or d2-test score, respectively.

For the analysis of performance and confidence at the trial level, data were analysed using linear and generalised linear mixed effects models. We always used the between-subject factor age as a predictor. The within-subject factor of interest was either accuracy (error, correct) or (pooled) confidence (3 levels). We fitted random intercepts for participants and, if possible, random slopes by participant for the within-subject factor of interest. For the outcome variables of RT, confidence and response caution, we fitted linear mixed models, for which *F* statistics are reported and degrees of freedom were estimated by Satterthwaite’s approximation, and for accuracy we fitted generalised linear mixed models, for which X^2 statistics are reported. Model structures and coefficients are reported in the supplementary material S1.

Significant effects of confidence were followed up by pairwise comparisons across rating levels using paired-samples *t*-tests for linear mixed models and *Z*-tests for generalised linear mixed models. Significant interactions were followed up by (generalised) linear mixed regressions, separately for each level of a given within-subject factor to assess potential effects of age. We decided on these follow-up tests because our main interest was in the differential relations between accuracy, confidence, and behaviour across the lifespan rather than between the levels. Post-hoc test results were compared against Holm corrected significance thresholds to account for multiple comparisons.

Analyses were run in MATLAB R2019a and R (version 4.0.5; R Core Team, 2021) using the lme4 package (version 1.1; Bates et al., 2015).

2.6. Electroencephalographic data analysis

One participant had to be removed from electroencephalographic analyses, because noisy EEG data led to the exclusion of more than half of the trials. Data were response-locked and analysed at the single trial level. We first computed the grand-average for all participants, separately for errors and correct responses. The latency of the grand-average peak amplitude served as the time point around which individual mean amplitudes were extracted from the signal (± 50 ms). This was done to obtain meaningful time windows for statistical analyses, because data of single trials is too noisy to identify a meaningful peak (Clayson et al., 2013). On each trial, the $N_{e/c}$ local peak amplitudes were extracted from the response-locked data from the interval 0 to 150 ms following the response at FCz, and the $P_{e/c}$ local peak amplitudes were extracted from the interval 150 to 350 ms at Cz. This was based on visual inspection of the local maxima of the grand-average scalp topographies as well as previous literature (Falkenstein et al., 2000; Siswandari et al., 2019).

For statistical analyses of ERP amplitudes, we fitted the same linear mixed effects regression models as for the behavioural data. They included fixed effects of age and the within-subject factor accuracy (error, correct) for all trials combined (see supplementary material S3 for the analysis with the within-subject factor confidence for all trials) or confidence (3 levels) for separate analysis of errors and correct responses, a random intercept for each participant, and a random slope of the within subject factor by participant, if possible. The models were fitted to the CSD-transformed single trial mean ERP amplitudes of the $N_{e/c}$ and $P_{e/c}$. Model structures and coefficients are reported in the supplementary material S2.

3. Results

For brevity, only significant effects in the mixed effects regression analyses and relevant follow-up tests are reported in this section. For results of all tests as well as Bayesian analyses of relevant null effects, please refer to the supplementary material S1, S2 and S6.

3.1. Behavioural results

3.1.1. Attention

The average score for sustained attention and processing speed as assessed by the d2-test was 178.5 ± 5.6 ($M \pm SEM$) and showed the typical decline for older adults, as reflected in a significant prediction of the test scores by age [$F(1,63) = 27.819$, $p < 0.001$; $\beta = -1.536$, $SE = 0.291$].

3.1.2. Distribution of confidence ratings

In a first step, we were interested in how the confidence ratings were distributed across the four confidence levels across the lifespan (Fig. 2). For this, we ran two Friedman ANOVAs for dependent

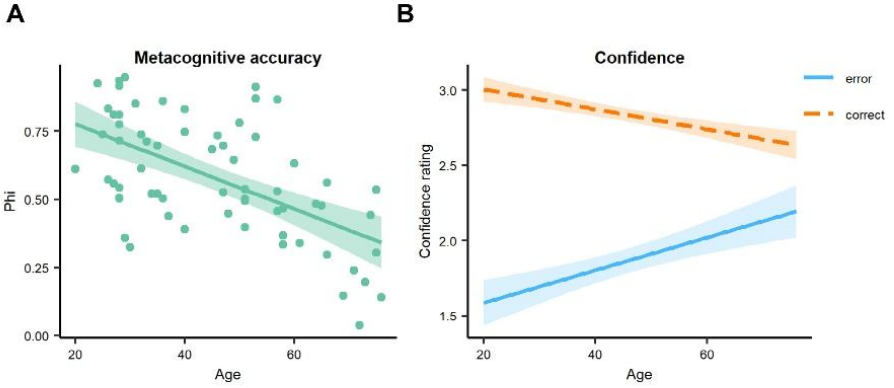


Fig. 3. Metacognition across the lifespan. (A) Metacognitive accuracy (Phi) decreased with age. Dots represent means of individual participants. (B) Confidence ratings for errors and correct trials were significantly predicted by age (in years). With increasing age, confidence was reduced for correct responses and increased for errors.

measures for the proportion for each rating category, separately for errors and correct responses.

The ANOVA for errors showed that the proportion differed between confidence levels [$X^2(3) = 78.029$, $p < 0.001$; Fig. 2A]. On average, most errors were rated as 'surely wrong' (42.8 %) and least errors as 'maybe wrong' (7.3%). Follow-up linear regressions on age-related differences for each rating category showed that the proportion of 'maybe correct' ratings was increased with higher age [$F(1,63) = 15.973$, $p < 0.001$; $\beta = 0.005$, $SE = 0.001$], whereas the ratio of 'surely wrong' ratings was decreased [$F(1,63) = 26.276$, $p < 0.001$; $\beta = -0.008$, $SE = 0.002$].

For correct responses, the ANOVA also revealed a main effect of confidence [$X^2(3) = 167.472$, $p < 0.001$]. Correct responses were most often rated as 'surely correct' (84.2 %) and least often as 'surely wrong' (0.7 %). Again, linear regression analyses on age-related differences showed that the proportion of 'maybe correct' ratings was increased with higher age [$F(1,63) = 24.653$, $p < 0.001$; $\beta = 0.006$, $SE = 0.001$], and the proportion of 'surely correct' ratings was decreased with age [$F(1,63) = 24.815$, $p < 0.001$; $\beta = -0.006$, $SE = 0.001$; Fig. 2B].

As mentioned above, to ensure a sufficient number of trials for each level of confidence for each participant, we combined 'maybe wrong' and 'maybe correct' ratings into one category representing 'unsure' responses. Thus, for all following behavioural analyses including the factor confidence, the reported analyses use three confidence levels.

3.1.3. *Error rate (ER)*

The average error rate was 15.6 ± 1.6%, and the mixed effects regression model testing for effects of confidence and age on error rate showed that the error rate significantly increased with higher age [$X^2(1) = 4.704$, $p = 0.030$]. The analysis further showed an effect of confidence on error rate [$X^2(2) = 2200.020$, $p < 0.001$]. The error rate decreased across confidence levels from 94.0 ± 0.7% on trials rated as 'surely wrong' to 67.3 ± 0.8% on trials rated as 'unsure' and 6.6 ± 0.2% on trials rated as 'surely correct'. Pairwise comparisons indicated that all comparisons were statistically significant (all $p < 0.001$). Thus, on average, participants' confidence reflected their performance well (which further supports the notion that the current study's confidence scale was a meaningful assessment tool). Furthermore, the regression analysis revealed a significant interaction between confidence and age [$X^2(2) = 168.125$, $p < 0.001$]. In subsequent mixed effects regression analyses for each level of confidence, error rates only significantly increased with higher age for the 'surely correct' confidence level [$X^2(1) = 37.664$, $p < 0.001$].

3.1.4. *Response time (RT)*

A mixed effects regression model predicting RT and testing for the effects of accuracy and age showed a significant effect of accuracy [$F(1,61.6) = 5.572$, $p = 0.021$] with on average slower errors (752.3 ± 3.8 ms) than correct responses (716.5 ± 1.3 ms). Moreover, the model revealed the expected slowing with age [$F(1,62.9) = 17.358$, $p < 0.001$], which did not significantly differ between errors and correct responses.

The mixed effects regression with the within-subject factor confidence similarly revealed an age-related slowing [$F(1,63.7) = 13.305$, $p < 0.001$; Fig. 4A]. Moreover, the analysis revealed an effect of confidence [$F(2,61.8) = 27.291$, $p < 0.001$] and a significant interaction between confidence and age [$F(2,56.6) = 5.187$, $p = 0.009$]. Pairwise comparisons indicated that all pairs were statistically significantly different (all $p < 0.010$), with trials associated with the 'unsure' confidence level (815.6 ± 6.7 ms) being considerably slower than trials rated as 'surely correct' (702.1 ± 1.3 ms) or 'surely wrong' (736.6 ± 6.7 ms). Furthermore, trials were significantly slower with older age for the extreme ratings ['surely wrong': $F(1,68.4) = 13.592$, $p < 0.001$; 'surely correct': $F(1.62.4) = 18.358$, $p < 0.001$], but not for 'unsure' ratings.

In short, RT was associated with confidence, such that high certainty (i.e. 'surely correct/wrong') was associated with the fastest responses, and this confidence-related modulation of RT decreased with higher age.

3.1.5. *Confidence*

A linear mixed effects regression model predicting confidence (coded from 1 to 3) across all trials revealed a significant effect of accuracy [i.e. error vs. correct trials; $F(1,63.4) = 162.928$, $p < 0.001$] and a significant interaction between accuracy and age [$F(1,62.4) = 37.361$, $p < 0.001$], but no significant effect of age. The average confidence rating was lower for errors (1.991 ± 0.014) compared to correct responses (2.867 ± 0.003). Follow-up regression analyses predicting confidence as a function of age for errors and correct responses separately revealed that confidence increased with age for errors [$F(1,60.1) = 17.977$, $p < 0.001$], while for correct responses it decreased [$F(1,62.1) = 23.816$, $p < 0.001$; Fig. 3B].

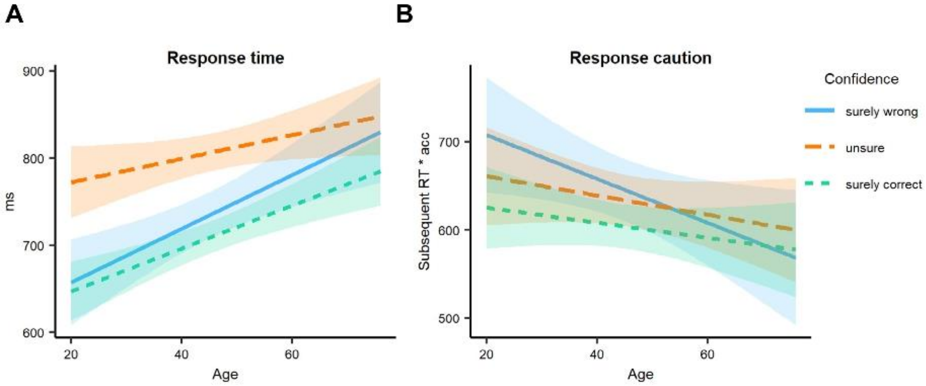


Fig. 4. Modulation of response time (RT; A) and response caution (B) by confidence and age (in years). (A) Trials rated as 'unsure' showed slower RTs than trials associated with any of the 'surely' rating categories, and this difference was smaller with increasing age. (B) Adaptation of response caution depending on previous trial confidence rating. Response caution was computed as the product of the accuracy and RT of subsequent trials. Across the lifespan, participants responded less cautiously after higher confidence ratings.

3.1.6. *Metacognitive accuracy (Phi)*

Phi had a mean of 0.579 ± 0.027 across the entire sample and was significantly predicted by age at the group level [$F(1,63) = 32.206$, $p < 0.001$; $\beta = -0.008$, $SE = 0.001$], indicating a decrease of metacognitive accuracy with age (Fig. 3A). Moreover, a multiple linear regression including the additional factor of error rate did not show a significant interaction with age ($p = 0.535$), suggesting that the association between metacognitive accuracy and age was not affected by decreased task performance in older adults. Similarly, a multiple linear regression including the additional factor of d2-test scores (which provide a task-independent measure of attention) suggested that the decrease in Phi with age was also independent of an age-related reduction in attentional capacity (interaction: $p = 0.091$).

3.1.7. *Behavioural adaptation*

To investigate the effect of accuracy and confidence in a given trial on the behaviour in the following trial, we computed response caution as the product of accuracy (coded as 0 and 1) and RT in the subsequent trial. The mixed effects regression with the within-subject factor accuracy (referring to the previous trial) revealed a significant effect of accuracy [$F(1,55.9) = 12.366$, $p < 0.001$] and an interaction between accuracy and age [$F(1,43.9) = 6.709$, $p = 0.013$], but no significant effect of age. Follow-up regression analyses for the subsets of errors or correct responses showed a nominal decrease in response caution with age for errors, but neither this nor the effect of age for correct responses was significant. Thus, these findings indicate that participants were on average more cautious after errors than after correct responses, and this effect did not significantly vary across age.

Next, we examined whether the response caution in the subsequent trial could also be predicted by the confidence rating in the preceding trial. As shown above, confidence and accuracy are strongly related; however, a significant modulation by confidence could also indicate that this internal confidence signal drives behavioural adaptations. The mixed effects regression on response caution with the within-subject factor confidence (referring to the previous trial) indeed revealed an effect of confidence [$F(2,54.9) = 7.306$, $p = 0.002$], but again, no effect of age and also no significant interaction (Fig. 4B). Pairwise comparisons between the confidence levels showed that the response caution after trials rated as 'surely correct' was significantly lower compared to trials rated as 'unsure' or as 'surely wrong'.

To summarise the effects of ageing on behaviour, we found the expected age-related general increase in error rates and response times, accompanied by a decrease in metacognitive ability, which was mainly reflected in reduced use of confidence ratings at the extreme ends of the scale but more indications of being unsure. Response caution, on the other hand, was not affected by ageing. Caution increased after errors compared to correct responses, and was notably specifically modulated by previous trial confidence. With higher confidence, the response caution in the subsequent trial decreased.

3.2. *Electrophysiological results*

3.2.1. *$N_{e/c}$ amplitudes*

The mean amplitude of the $N_{e/c}$ was significantly larger for errors compared to correct responses, as reflected in an effect of accuracy in the mixed effects regression predicting the $N_{e/c}$ as a function of accuracy and age [$F(1,38.6) = 9.054$, $p = 0.005$; Fig. 5A]. There was no main effect of age, but a significant interaction [$F(1,31.8) = 5.472$, $p = 0.026$] as the amplitude of the N_e [$F(1,55.4) = 5.030$, $p = 0.029$] but not the N_c was smaller with higher age.

For the analysis of confidence, we fitted separate linear mixed effects models to the N_e amplitude for errors and to the N_c amplitude for correct responses, with confidence as the within-subject factor and age as the between-subject factor. The regression analysis for errors showed effects of age [$F(1,57.4) = 4.068$, $p = 0.048$], confidence [$F(2,2706.4) = 4.007$, $p = 0.018$], and a significant interaction between age and confidence [$F(2,2731.5) = 3.662$, $p = 0.026$; Fig. 6A]. Pairwise comparisons between the confidence levels indicated a significant difference between errors rated as 'surely wrong' and 'surely correct', and follow-up mixed effects regressions showed that specifically the N_e amplitudes of low confidence errors (i.e. rated as 'surely wrong') was decreased with older age [$F(1,58.1) = 9.735$, $p = 0.003$].

The regression analysis for correct responses with the within-subject factor confidence yielded no significant effects (Fig. 6B).

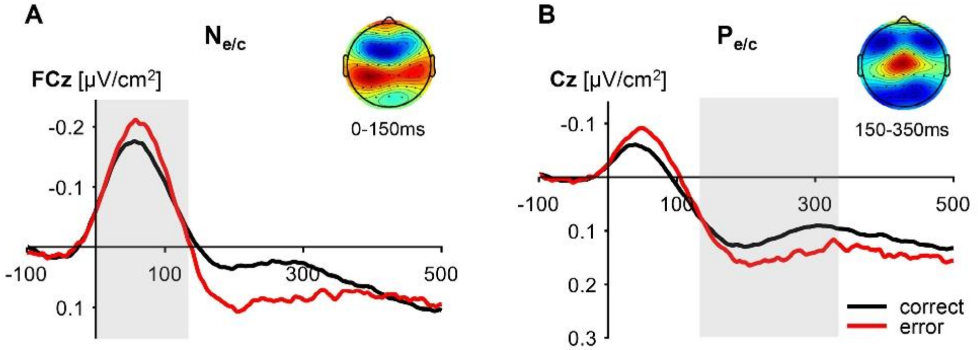


Fig. 5. Response-locked event-related potentials for errors and correct responses and topographical maps of errors after current source density transformation. (A) $N_{e/c}$ is computed at electrode FCz and (B) $P_{e/c}$ at electrode Cz. Errors are shown in red, correct trials in black. Scalp topographies depict the mean activity for all error trials averaged across the time windows for the N_e (0-150 ms) and the P_e (150-350 ms). Grey squares indicate time windows for the identification of peak amplitudes, which served to compute the adaptive mean amplitudes.

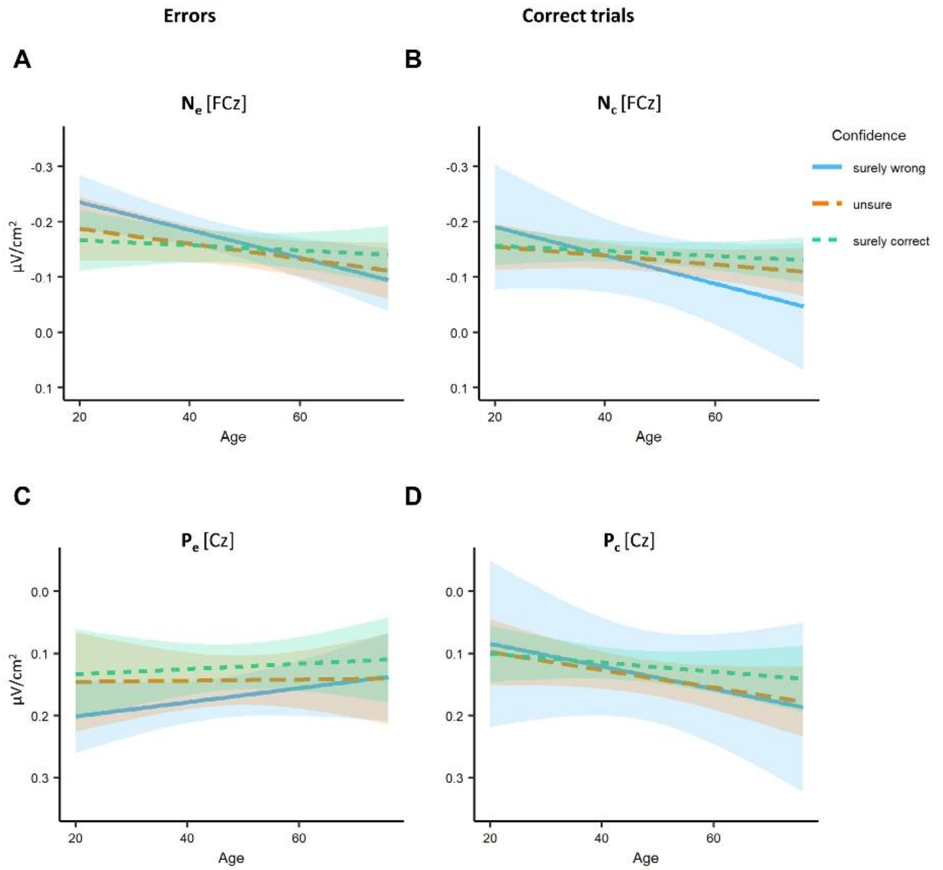


Fig. 6. Regression of response-locked event-related potentials on age (in years) by confidence, separately for errors and correct responses after current source density transformation. Errors are shown in the left panel, correct trials in the right panel. The N_e (A) and N_c (B) are shown at electrode FCz, and P_e (C) and P_c (D) are shown at electrode Cz. For errors, the amplitudes increased with lower confidence, while for correct responses, they were not modulated by confidence. Age predicted a decrease in N_e amplitude of 'surely wrong' errors.

3.2.2. $P_{e/c}$ amplitudes

The mixed effects regression on the $P_{e/c}$ amplitude with the within-subject factor accuracy revealed a significant effect of accuracy with larger amplitudes for errors compared to correct responses [$F(1,55.3) = 10.378$, $p = 0.002$; Fig. 5B]. There was no effect of age, but a significant interaction between accuracy and age [$F(1,49.2) = 6.443$, $p = 0.014$]. However, in follow-up regression analyses, no significant associations were found for errors or correct responses.

Next, responses were again split by their accuracy, and separate linear mixed effects models were fitted to the P_e and P_c amplitudes, respectively, with the within-subject factor confidence. Neither the analysis for errors nor the analysis for correct responses yielded any significant effects on the P_e and P_c amplitudes (Fig. 6C and D).

However, due to previous evidence suggesting a strong relation between $P_{e/c}$ amplitude and error detection or confidence ratings (Boldt and Yeung, 2015; Nieuwenhuis et al., 2001), we were specifically interested in the modulation of the P_e by confidence. To replicate previous findings, we fitted an additional, exploratory mixed effects model to the P_e amplitudes, including only the factor of confidence. The analysis revealed a significant, albeit small difference in P_e amplitude between errors rated as ‘surely wrong’ and errors rated as ‘surely correct’ [$F(2,2723.6) = 6.627$, $p = 0.001$], as confirmed in follow-up multiple comparisons between confidence levels ($t = 3.617$, $p < 0.001$). This exploratory analysis implies that the P_e was modulated by confidence when assessed independent of age.

4. Discussion

We conducted a complex four-choice flanker task with adult participants covering an age range from 20 to 76 years, allowing us to investigate confidence and metacognitive accuracy as well as neural indices thereof across the lifespan. We found that error rates and response times (RT) increased with age. Metacognitive accuracy, quantified as Phi, gradually decreased across the lifespan and was characterised by a differential use of confidence ratings. In contrast, we did not find differences between younger and older adults in the ability to adapt behaviour in accordance with reported confidence. As expected, the $N_{e/c}$ and $P_{e/c}$ amplitudes declined with higher confidence in having made a correct response, which was specifically observed for trials with response errors. While the N_e amplitude was smaller with older age whenever participants were sure they made an error, the variation in the P_e amplitude with reported confidence was surprisingly not affected by ageing. In the following, we will first discuss potential processes underlying age-related differences in metacognitive accuracy and their relation to task performance and confidence, before comparing the pattern we observed at the behavioural level to the patterns we observed in the ERPs. Finally, we argue that older adults' preserved ability to adapt their behaviour to their perceived confidence could be related to the $P_{e/c}$ amplitude.

4.1. Differential use of confidence scale as a marker of age-related metacognitive decline

In the present study, metacognitive accuracy (Phi) was reduced with increasing age. This is consistent with the findings of Palmer et al. (2014) who used a metacognitive efficiency measure, which further considered the individual performance in their perceptual discrimination task. As this measure was not directly applicable in our four-choice flanker task, we confirmed (by calculating multiple linear regressions taking into account the error rate and the d2-test score) that the observed decline in metacognitive accuracy was not merely a reflection of general age-related performance or attention deficits (d2-test; see also Larson & Clayson, 2011). Our results, therefore, show that Palmer et al.'s (2014) findings also hold for a more complex, speeded decision task, which was not based on stimulus ambiguity.

The question remains as to how the age-related differences in confidence emerge. Given the nature of Phi, a smaller value could either indicate more undetected errors or correct responses rated as being incorrect, or a generally higher uncertainty (i.e. rating all correct responses as ‘maybe correct’ will result in a lower Phi value than rating the same number of correct responses as ‘surely correct’). Indeed, we observed that older adults used the extreme ends of the confidence scale considerably less often than younger adults.

For errors, this pattern resulted in a higher mean confidence with age. This disproportional rise in reported confidence has similarly been shown in error detection studies, indicated by a lower error detection rate in older adults (Harty et al., 2017,2013; Niessen et al., 2017). For correct decisions, we observed a *lower* mean confidence due to the tendency of the older adults to use the middle of the confidence scale, whereas previous studies rather reported an *over*-confidence in older age (Dodson et al., 2007; Hansson et al., 2008; Ross et al., 2012).

Interestingly, participants in our study responded slowest in case of uncertainty, i.e. ‘unsure’ ratings. In contrast, studies on decision confidence typically report increasing RT with decreasing confidence (Kiani et al., 2014; Rahnev et al., 2020; Weidemann and Kahana, 2016). Most of these studies specifically measured confidence in having made a correct decision (i.e. the lowest confidence indicates guessing, while in our study it indicates high certainty in being incorrect), and typical paradigms in these studies are two-choice signal detection tasks in which the degree of sensory evidence, for instance, perceptual discriminability is manipulated (Kiani et al., 2014; Moran et al., 2015; Rollwage et al., 2020). In our task, we ensured (using a designated colour discrimination test) that all stimuli were perceptually discriminable without time pressure, and our data showed no signs of age-related differences in stimulus processing (even though it remains possible that slight impairments in colour perception, or other untested factors such as attention, working memory, etc., might have contributed to the age-related slowing we observed; see supplementary material S4). Instead, potential sources for errors could be, for instance, stimulus conflict caused by the flankers and the similarity of the stimulus colours, or difficulties in remembering the stimulus response mapping. Using a comparable paradigm, Stahl et al. (2020) found slow errors to be associated with lower confidence than fast, impulsive errors and inferred that those error types should predominantly be caused by weak stimulus-response representations (i.e. due to weak memory traces).

As such conclusions could not be drawn from classical error processing studies requiring only a binary error detection rating, our findings provide an important link between those and decision confidence studies. In a typical error processing paradigm that posed higher demands on the older adults (as indicated, for instance, by higher error rates), our results could be interpreted as their impaired metacognitive evaluation (assessed via confidence ratings) being partly related to more frequent memory-related errors, which appear to be more challenging to assess consciously (Maier and Steinhauser, 2017; Stahl et al., 2020).

4.2. Neural correlate of confidence is stable across age

The $P_{e/c}$ is an established marker of metacognition, reflecting variations in subjective error awareness and decision confidence (Boldt and Yeung, 2015; Nieuwenhuis et al., 2001). In the present study, the $P_{e/c}$ showed the well-known accuracy effect of larger amplitudes for errors than correct responses. Moreover, we could replicate prior findings of the P_e increasing with decreasing confidence, - for the first time - for a very broad age range (Boldt and Yeung, 2015; Rausch et al., 2019). This also replicates findings from error detection studies showing increased P_e amplitudes for detected compared to undetected errors (Endrass et al., 2012a; Nieuwenhuis et al., 2001).

The main interest of our study was to investigate the modulation of the $P_{e/c}$ by metacognition in the context of healthy ageing. Remarkably, the $P_{e/c}$ amplitude did not show an overall reduction with age, nor a differential modulation by confidence across the lifespan, suggesting that the accumulation of error evidence was well preserved in older age. This is contrary to the error detection literature (Harty et al., 2017; Niessen et al., 2017). Since these studies did not assess confidence on multiple levels, participants did not have the chance to express uncertainty. Assuming more 'unsure' cases with older age, their observed age-related decrease in P_e amplitude for detected errors might thus be confounded, as higher uncertainty was generally associated with reduced P_e amplitudes (Boldt and Yeung, 2015). Following this logic, it is also possible to explain the lack of a significant age-related modulation of the $P_{e/c}$ amplitude in the present study: If older adults' internal threshold for rating an error as 'surely wrong' was generally raised, the errors that *were* rated as 'surely wrong' should be trials with particularly high P_e amplitudes, as they were absolutely sure of having committed an error. As a result, a putative age-related decrease in the P_e amplitude of low confidence errors could be masked in our data, because the same reported rating levels might reflect a different sense of confidence for younger and older adults. Thus, the current pattern of results suggests that the P_e amplitude does *not* serve as a direct index of metacognitive accuracy across participants, but rather reflects the degree of confidence, irrespective of objective performance (Di Gregorio et al., 2018; Larson and Clayson, 2011; Pouget et al., 2016; Stahl et al., 2020).

4.3. Impaired neural processing of conflict modulates metacognitive decline

The marked behavioural decline in older adults' metacognitive accuracy was not mirrored in age-related variations of the $P_{e/c}$ amplitude, but rather in a differential modulation of the N_e across the lifespan. The modelling results revealed that the N_e amplitude was also affected by the interaction between confidence and age. With older age, the N_e declined for all errors in which high conflict was perceived. In other words, only the N_e amplitude of errors which were rated as 'surely wrong' varied in amplitude across the lifespan. As the $N_{e/c}$ is sensitive to conflict between the given and the actual correct response, older adults seemed to having had difficulties internally representing the correct response in highly conflicting situation (Yeung et al., 2004). Notably, this effect was error-specific, that is, we cannot draw conclusions about internal processes for correct responses, as the N_c amplitude did not show a relation to confidence that could have varied with age.

We suggest that the reduced N_e amplitude of low confidence errors with higher age could be related to the observed decrease in metacognitive accuracy in our flanker task. If older adults did not perceive high conflict due to difficulties in forming an accurate internal representation of the correct response, this information was necessarily missing for the metacognitive evaluation. Thus, the impaired neural integration of conflict detection and confidence could have led to the observed behavioural difficulties matching confidence ratings and objective accuracy.

4.4. Adults of all ages base future behaviour on subjective confidence

Ultimately, proper metacognitive evaluation should improve behaviour. Interestingly, response caution was not only enhanced after errors, but we also found evidence that it was modulated by the reported confidence in the preceding trial. Given that the participants did not receive any external feedback about the accuracy of their response (as it is often the case in real-life decisions), it seems plausible that they used their best available estimate, i.e. the subjective sense of confidence, to regulate subsequent behaviour (Desender et al., 2019a). Specifically, low confidence (reflecting a belief that an error had been committed) or uncertainty about a decision were associated with higher response caution in the subsequent trial. Possibly, participants sought more evidence before committing to their next decision, leading to slower but more accurate responses (Desender et al., 2019a, 2019b).

Translating our findings to error detection studies, the increase in response caution with lower previous trial confidence converges with findings of error detection studies reporting increased slowing (i.e. a sign of behavioural adaptation) after detected compared to undetected errors (Nieuwenhuis et al., 2001; Stahl et al., 2020; Wessel et al., 2018; for a review on post-error adjustments see Danielmeier & Ullsperger, 2011).

Notably, response caution was similarly affected by accuracy and confidence across the lifespan. Thus, while metacognitive accuracy was reduced in older age, a neural correlate of error confidence magnitude, the P_e amplitude, and the behavioural adaptations relative to the reported confidence were consistent across the lifespan. This suggests that it is the perceived confidence that shapes future behaviour, irrespective of metacognitive accuracy: Despite their failure in matching confidence to task performance, older adults seem to be equally able to use internal states of confidence to change future behaviour adaptively.

4.5. Limitations and implications

One limitation of the present study is the number of participants retained for the analyses. When designing the experiment, we tried to find an optimal balance between task difficulty, feasibility for all ages, and gaining many trials while ensuring that especially older adults were not exhausted at the end of the experiment. However, the combination of a substantial number of response alternatives, time pressure, and discriminability of stimuli was demanding, leading to an undesirably large number of participants to be excluded from the analyses (17 of the initial 82 participants).

A second shortcoming is the confined number of trials available for analysis after defining conditions of interest. Due to an unforeseen highly skewed use of the confidence scale, it was impossible to apply a factorial design while retaining four distinct confidence levels. In particular for correct trials, the variance in confidence ratings was low, which is a common problem in metacognition research (for a review, see Wessel, 2012). However, the application of linear mixed effects modelling provided us with a powerful tool that can account for varying trial numbers across participants and importantly, the multi-level structure of our data.

Nevertheless, our findings provide important insights into ageing effects on metacognition, integrating approaches from error detection and decision confidence research. In contrast to the metacognitive evaluation itself, the effect of confidence on subsequently adapting response caution was well preserved in older adults. Thus, training the metacognitive evaluation of fundamental decisions in older adults might constitute a promising endeavour (and has been shown to work for mathematical problem solving [Pennequin et al., 2010]).

5. Conclusion

The study of error detection and confidence in the context of healthy ageing have advanced largely in parallel. Our study demonstrates that confidence shapes our behavioural and neural processing of decisions and should be considered to investigate age-related effects on error processing and metacognitive abilities. Interestingly, the N_e, but not the P_e amplitude was differentially modulated by confidence across the lifespan, suggesting that the decreasing accuracy of metacognitive judgements with older age might be related to impaired integration of neural correlates of conflict detection and decision confidence.

Disclosure statement

The authors declare no conflict of interest.

Verification for the manuscript

"Neural correlates of metacognition across the adult lifespan"

1 All authors declare that they have no potential conflicts of interest to disclose.
2 This work has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation – Project-ID 431549029 – SFB 1451; GRF and PHW), and the Australian Research Council (Discovery Project Grant; DP160103353; SB).
3 The data and analyses contained in this manuscript have not been submitted elsewhere and will not be submitted elsewhere while under consideration at *Neurobiology of Aging*.
4 The study was approved by the ethics committee of the German Psychological Society (DGPs) and conformed to the Declaration of Helsinki.
5 All authors have reviewed the contents of the manuscript being submitted, approved its contents and validated the accuracy of the data.

CRediT authorship contribution statement

Helen Overhoff: Conceptualization, Methodology, Software, Investigation, Data curation, Formal analysis, Writing – original draft, Visualization, Writing – review & editing. **Yiu Hong Ko:** Conceptualization, Writing – review & editing. **Daniel Feuerriegel:** Methodology, Writing – review & editing. **Gereon R. Fink:** Resources, Writing – review & editing. **Jutta Stahl:** Conceptualization, Methodology, Software, Writing – review & editing, Supervision. **Peter H. Weiss:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision. **Stefan Bode:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision. **Eva Niessen:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Supervision.

Acknowledgements

We thank all colleagues from the Institute of Neuroscience and Medicine (INM-3), Cognitive Neuroscience, and the Decision Neuroscience Lab at the Melbourne School of Psychological Sciences for valuable discussions and their support.

This work has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation – Project-ID 431549029 – SFB 1451; GRF and PHW), and the Australian Research Council (Discovery Project Grant; DP160103353; SB).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neurobiolaging.2021.08.001.

References

Bates, D., Mächler, M., Boker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using {lme4}. J. Stat. Softw. 67, 1–48. doi:10.18637/jss.v067.i01.

Boldt, A., Yeung, N., 2015. Shared neural markers of decision confidence and error detection. J. Neurosci. 35, 3478–3484. doi:10.1523/JNEUROSCI.0797-14.2015.

Brickenkamp, R., 2002. The d2 Test of attention, 1st ed.9th ed. Verlag für Psychologie Hogrefe, Goettingen.

Castel, A.D., Middlebrooks, C.D., McGillivray, S., 2016. Monitoring Memory in Old Age: Impaired, Spared, and Aware. In: Dunlosky, J., Tauber, S.K. (Eds.), The Oxford Handbook of Metamemory. Oxford University Press, pp. 519–534. doi:10.1093/oxfordhb/9780199336746.013.3.

Chua, E., Schacter, D.L., Sperling, R., 2009. Neural basis for recognition confidence. Psychol. Aging 24, 139–153. doi:10.1037/a0014029.Neural.

Clawson, A., Clayson, P.E., Keith, C.M., Catron, C., Larson, M.J., 2017. Conflict and performance monitoring throughout the lifespan: An event-related potential (ERP) and temporospatial component analysis. Biol. Psychol. 124, 87–99. doi:10.1016/j.biopsycho.2017.01.012.

Clayson, P.E., Baldwin, S.A., Larson, M.J., 2013. How does noise affect amplitude and latency measurement of event-related potentials (ERPs)? A methodological critique and simulation study. Psychophysiology 50, 174–186. doi:10.1111/psyp.12001.

Danielmeier, C., Ullsperger, M., 2011. Post-error adjustments. Front. Psychol. 2, 1–10.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. 134, 9–21.

Desender, K., Boldt, A., Verguts, T., Donner, T.H., 2019a. Confidence predicts speed-accuracy tradeoff for subsequent decisions. Elife doi:10.1101/466730.

Desender, K., Murphy, P.R., Boldt, A., Verguts, T., Yeung, N., 2019b. A Postdecisional Neural Marker of Confidence Predicts Information-Seeking in Decision-Making. J. Neurosci. 39, 3309–3319. doi:10.1101/433276.

Di Gregorio, F., Maier, M.E., Steinhauser, M., 2018. Errors can elicit an error positivity in the absence of an error negativity: Evidence for independent systems of human error monitoring. Neuroimage 172, 427–436. doi:10.1016/j.neuroimage.2018.01.081.

Dodson, C.S., Bawa, S., Krueger, L.E., 2007. Aging, metamemory, and high-confidence errors: A misrecollection account. Psychol. Aging 22, 122–133. doi:10.1037/0882-7974.22.1.122.

Dully, J., McGovern, D.P., O'Connell, R.G., 2018. The impact of natural aging on computational and neural indices of perceptual decision making: A review. Behav. Brain Res. 355, 48–55. doi:10.1016/j.bbr.2018.02.001.

Endrass, T., Klawohn, J., Preuss, J., Kathmann, N., 2012a. Temporospatial dissociation of Pe subcomponents for perceived and unperceived errors. Front. Hum. Neurosci. 6, 1–10. doi:10.3389/fnhum.2012.00178.

Endrass, T., Schreiber, M., Kathmann, N., 2012b. Speeding up older adults: Age-effects on error processing in speed and accuracy conditions. Biol. Psychol. 89, 426–432. doi:10.1016/j.biopsycho.2011.12.005.

Eriksen, B.A., Eriksen, C.W., 1974. Effects of noise letters upon the identification of a target letter in a nonsearch task. Percept. Psychophys. 16, 143–149.

Falkenstein, M., Hohnsbein, J., Hoormann, J., Blanke, L., 1991. Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. Electroencephalogr. Clin. Neurophysiol. 78, 447–455. doi:10.1016/0013-4694(91)90062-9.

Falkenstein, M., Hoormann, J., Christ, S., Hohnsbein, J., 2000. ERP components on reaction errors and their functional significance: A tutorial. Biol. Psychol. 51, 87–107. doi:10.1016/S0301-0511(99)00031-9.

Falkenstein, M., Hoormann, J., Hohnsbein, J., 2001. Changes of error-related ERPs with age. Exp. Brain Res 138, 258–262. doi:10.1007/s002210100712.

Fitzgerald, L.M., Arvaneh, M., Dockree, P.M., 2017. Domain-specific and domain-general processes underlying metacognitive judgments. Conscious. Cogn. 49, 264–277. doi:10.1016/j.concog.2017.01.011.

Fleming, S.M., Dolan, R.J., 2012. The neural basis of metacognitive ability. royalsocietypublishing.org 367, 1338–1349. doi:10.1098/rstb.2011.0417.

Fleming, S.M., Dolan, R.J., Frith, C.D., 2012. Metacognition: Computation, biology and function. Philos. Trans. R. Soc. B Biol. Sci. 367, 1280–1286. doi:10.1098/rstb.2012.0021.

Fleming, S.M., Frith, C.D., 2014. The Cognitive Neuroscience of Metacognition, The Cognitive Neuroscience of Metacognition. Springer doi:10.1007/978-3-642-45190-4.

Fleming, S.M., Lau, H.C., 2014. How to measure metacognition. Front. Hum. Neurosci. 8, 1–9. doi:10.3389/fnhum.2014.00443.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. 12, 189–198. doi:10.1016/0022-3956(75)90026-6.

Galvin, S.J., Podd, J.V., Drga, V., Whitmore, J., 2003. Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. Psychon. Bull. Rev. 10, 843–876. doi:10.3758/BF03196546.

Groom, M.J., Cragg, L., 2015. Differential modulation of the N2 and P3 event-related potentials by response conflict and inhibition. Brain Cogn 97, 1–9. doi:10.1016/j.bandc.2015.04.004.

Hansson, P., Rönnlund, M., Juslin, P., Nilsson, L.G., 2008. Adult Age Differences in the Realism of Confidence Judgments: Overconfidence, Format Dependence, and Cognitive Predictors. Psychol. Aging 23, 531–544. doi:10.1037/a0012782.

Harty, S., Murphy, P.R., Robertson, I.H., O’Connell, R.G., 2017. Parsing the neural signatures of reduced error detection in older age. Neuroimage 161, 43–55. doi:10.1016/j.neuroimage.2017.08.032.

Harty, S., O’Connell, R.G., Hester, R., Robertson, I.H., 2013. Older adults have diminished awareness of errors in the laboratory and daily life. Psychol. Aging 28, 1032–1041. doi:10.1037/a0033567.

Hautzinger, M., 1991. The Beck Depression Inventory in clinical practice. Nervenarzt 62, 689–696.

Hertzog, C., 2015. Aging and Metacognitive Control. The Oxford Handbook of Metamemory doi:10.1093/oxfordhb/9780199336746.013.31.

Hertzog, C., Hultsch, D., 2000. Metacognition in adulthood and old age. In: Craik, F.I.M., Salthouse, T.A. (Eds.), The Handbook of Aging and Cognition. Lawrence Erlbaum Associates Publishers, pp. 417–466.

Hoerold, D., Pender, N., Robertson, IH, 2013. mende. Neuropsychologia 51, 385–391.

Jasper, H., 1958. The ten twenty electrode system of the international federation. Electroencephalogr. Clin. Neurophysiol. 10, 371–375.

Kayser, J., Tenke, C.E., 2006. Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: II. Adequacy of low-density estimates. Clin. Neurophysiol. 117, 369–380. doi:10.1016/j.clinph.2005.08.033.

Kiani, R., Corthell, L., Shadlen, M.N., 2014. Choice certainty is informed by both evidence and decision time. Neuron 84, 1329–1342. doi:10.1016/j.neuron.2014.12.015.

Kornell, N., Son, L.K., Terrace, H.S., 2007. Transfer of metacognitive skills and hint seeking in monkeys: Research article. Psychol. Sci. 18, 64–71. doi:10.1111/j.1467-9280.2007.01850.x.

Korsch, M., Frühholz, S., Herrmann, M., 2016. Conflict-specific aging effects mainly manifest in early information processing stages-an ERP study with different conflict types. Front. Aging Neurosci. 8, 1–12. doi:10.3389/fnagi.2016.00053.

Larson, M.J., Clayson, P.E., 2011. The relationship between cognitive performance and electrophysiological indices of performance monitoring. Cogn. Affect. Behav. Neurosci. 11, 159–171. doi:10.3758/s13415-010-0018-6.

Larson, M.J., Clayson, P.E., Keith, C.M., Hunt, I.J., Hedges, D.W., Nielsen, B.L., Call, V.R.A., 2016. Cognitive control adjustments in healthy older and younger adults: Conflict adaptation, the error-related negativity (ERN), and evidence of generalized decline with age. Biol. Psychol. 115, 50–63. doi:10.1016/j.biopsycho.2016.01.008.

Larson, M.J., South, M., Clayson, P.E., 2011. Sex differences in error-related performance monitoring. Neuroreport 22, 44–48. doi:10.1097/WNR.0b013e3283427403.

Lopez-Calderon, J., Luck, S.J., 2014. ERPLAB: An open-source toolbox for the analysis of event-related potentials. Front. Hum. Neurosci. 8, 1–14. doi:10.3389/fnhum.2014.00213.

Lucci, G., Berchicci, M., Spinelli, D., Taddei, F., Di Russo, F., 2013. The Effects of Aging on Conflict Detection. PLoS One 8. doi:10.1371/journal.pone.0056566.

Maier, M.E., Steinhauser, M., 2017. Working memory load impairs the evaluation of behavioral errors in the medial frontal cortex. Psychophysiology 54, 1472–1482. doi:10.1111/psyp.12899.

Maier, M.E., Steinhauser, M., Hübner, R., 2008. Is the Error-related Negativity Amplitude Related to Error Detectability? Evidence from Effects of Different Error Types. J. Cogn. Neurosci. 20, 2263–2273.

Maniscalco, B., Lau, H., 2012. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. Conscious. Cogn. 21, 422–430. doi:10.1016/j.concog.2011.09.021.

Mattler, U., 2003. Delayed flanker effects on lateralized readiness potentials. Exp. Brain Res 151, 272–288. doi:10.1007/s00221-003-1486-5.

Mecacci, L., Righi, S., 2006. Cognitive failures, metacognitive beliefs and aging. Pers. Individ. Dif. 40, 1453–1459. doi:10.1016/j.paid.2005.11.022.

Moran, R., Teodorescu, A.R., Usher, M., 2015. Post choice information integration as a causal determinant of confidence: Novel data and a computational account. Cogn. Psychol. 78, 99–147. doi:10.1016/j.cogpsych.2015.01.002.

Murphy, P.R., Robertson, I.H., Harty, S.S., O’Connell, R.G., O’Connell, R.G., 2015. Neural evidence accumulation persists after choice to inform metacognitive judgments. Elife 4, 1–23. doi:10.7554/eLife.11946.

Nelson, T., 1984. A comparison of current measures of the accuracy of feeling-of-knowing predictions. Psychol. Bull. 95, 109–133.

Niessen, E., Fink, G.R., Hoffmann, H.M., Weiss, P.H., Stahl, J., 2017. Error detection across the adult lifespan: Electrophysiological evidence for age-related deficits. Neuroimage 152, 517–529. doi:10.1016/j.neuroimage.2017.03.015.

Nieuwenhuis, S., Richard Ridderinkhof, K., Blom, J., Band, G.P.H., Kok, A., 2001. Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. Psychophysiology 38, 752–760. doi:10.1017/S0048577201001111.

Oldfield, R.C., 1971. The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9, 97–113. doi:10.1016/0028-3932(71)90067-4.

Palmer, E.C., David, A.S., Fleming, S.M., 2014. Effects of age on metacognitive efficiency. Conscious. Cogn. 28, 151–160. doi:10.1016/j.concog.2014.06.007.

Pennequin, V., Sorel, O., Mainguy, M., 2010. Metacognition, executive functions and aging: The effect of training in the use of metacognitive skills to solve mathematical word problems. J. Adult Dev. 17, 168–176. doi:10.1007/s10804-010-9098-3.

Perrin, F., Pernier, J., Bertrand, O., Echallier, J.F., 1989. Spherical splines for scalp potential and current density mapping. Electroencephalogr. Clin. Neurophysiol. 72, 184–187. doi:10.1016/0013-4694(89)90180-6.

Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. 118, 2128–2148. doi:10.1016/j.clinph.2007.04.019.

Pouget, A., Drugowitsch, J., Kepecs, A., 2016. Confidence and certainty: distinct probabilistic quantities for different goals. nature.com doi:10.1038/nn.4240.

Rabbitt, P., 1990. Age, IQ and awareness, and recall of errors. Ergonomics 33, 1291–1305. doi:10.1080/00140139008925333.

Rabbitt, P.M.A., 1966. Errors and error correction in choice-response tasks. Exp. Psychol. 71, 264–272. doi:10.1037/h0022853.

Rahnev, D., Desender, K., Lee, A.L.F., Adler, W.T., Aguilar-Lleyda, D., Akdoğan, B., Arbuzova, P., Atlas, L.Y., Balcı, F., Bang, J.W., Bègue, I., Birney, D.P., Brady, T.F., Calder-Travis, J., Chetverikov, A., Clark, T.K., Davranche, K., Denison, R.N., Dildine, T.C., Double, K.S., Duyan, Y.A., Faivre, N., Fallow, K., Filevich, E., Gajdos, T., Gallagher, R.M., de Gardelle, V., Gherman, S., Haddara, N., Hainguerlot, M., Hsu, T.Y., Hu, X., Iturrate, I., Jaquiery, M., Kantner, J., Koculak, M., Konishi, M., Koß, C., Kvam, P.D., Kwok, S.C., Lebreton, M., Lempert, K.M., Ming Lo, C., Luo, L., Maniscalco, B., Martin, A., Massoni, S., Matthews, J., Mazancieux, A., Merfeld, D.M., O'Hora, D., Palser, E.R., Paulewicz, B., Pereira, M., Peters, C., Philiastides, M.G., Pfuhl, G., Prieto, F., Rausch, M., Recht, S., Reyes, G., Rouault, M., Sackur, J., Sadeghi, S., Samaha, J., Seow, T.X.F., Shekhar, M., Sherman, M.T., Siedlecka, M., Skóra, Z., Song, C., Soto, D., Sun, S., van Boxtel, J.J.A., Wang, S., Weidemann, C.T., Weindel, G., Wierzchoń, M., Xu, X., Ye, Q., Yeon, J., Zou, F., Zylberberg, A., 2020. The Confidence Database. Nat. Hum. Behav. 4, 317–325. doi:10.1038/s41562-019-0813-1.

Rausch, M., Zehetleitner, M., Steinhauser, M., Maier, M.E., 2019. Cognitive modelling reveals distinct electrophysiological markers of decision confidence and error monitoring. bioRxiv, 860379 doi:10.1101/860379.

Rollwage, M., Loosen, A., Hauser, T.U., Moran, R., Dolan, R.J., Fleming, S.M., 2020. Confidence drives a neural confirmation bias. Nat. Commun. 11. doi:10.1038/s41467-020-16278-6.

Ross, L.A., Dodson, J., Edwards, J.D., Ackerman, M.L., Ball, K., 2012. Self-rated Driving and Driving Safety in Older Adults. Accid. Anal. Prev. 48, 523–527. doi:10.1038/jid.2014.371.

Ruitenberg, M.F.L., Abrahamse, E.L., De Kleine, E., Verwey, W.B., 2014. Post-error slowing in sequential action: An aging study. Front. Psychol. 5, 1–8. doi:10.3389/fpsyg.2014.00119.

Scheffers, M.K., Coles, M.G.H., 2000. Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. J. Exp. Psychol. Hum. Percept. Perform. 26, 141–151. doi:10.1037/0096-1523.26.1.141.

Schreiber, M., Pietschmann, M., Kathmann, N., Endrass, T., 2011. ERP correlates of performance monitoring in elderly. Brain Cogn 76, 131–139. doi:10.1016/j.bandc.2011.02.003.

Shekhar, M., Rahnev, D., 2020. The Nature of Metacognitive Inefficiency in Perceptual Decision Making. Psychol. Rev. 1–83. doi:10.1037/rev0000249.

Sim, J., Brown, F., O'Connell, R., Hester, R., 2020. Impaired error awareness in healthy older adults: an age group comparison study. Neurobiol. Aging. doi:10.1016/j.neurobiolaging.2020.08.001.

Siswandari, Y., Bode, S., Stahl, J., 2019. Performance monitoring beyond choice tasks: The time course of force execution monitoring investigated by event-related potentials and multivariate pattern analysis. Neuroimage 197, 544–556. doi:10.1016/j.neuroimage.2019.05.006.

Song, C., Kanai, R., Fleming, S.M., Weil, R.S., Schwarzkopf, D.S., Rees, G., 2011. Relating inter-individual differences in metacognitive performance on different perceptual tasks. Conscious. Cogn. 20, 1787–1792. doi:10.1016/j.concog.2010.12.011.

Stahl, J., Mattes, A., Hundrieser, M., Kummer, K., Mück, M., Niessen, E., Porth, E., Siswandari, Y., Wolters, P., Dummel, S., 2020. Neural correlates of error detection during complex response selection: Introduction of a novel eight-alternative response task. Biol. Psychol. 156. doi:10.1016/j.biopsycho.2020.107969.

Steinhauser, M., Yeung, N., 2010. Decision processes in human performance monitoring. J. Neurosci. 30, 15643–15653. doi:10.1523/JNEUROSCI.1899-10.2010.

Team, R.C., 2021. No Title.

Vidal, F., Burle, B., Bonnet, M., Grapperon, J., Hasbroucq, T., 2003. Error negativity on correct trials: a reexamination of available data. Biol. Psychol. 64, 265–282. doi:10.1016/S0301-0511(03)00097-8.

Weidemann, C.T., Kahana, M.J., 2016. Assessing recognition memory using confidence ratings and response times. R. Soc. Open Sci. 3, 150670. doi:10.1098/rsos.150670.

Wessel, J.R., 2012. Error awareness and the error-related negativity: evaluating the first decade of evidence. Front. Hum. Neurosci. 6, 1–16. doi:10.3389/fnhum.2012.00088.

Wessel, J.R., Dolan, K.A., Hollingworth, A., 2018. A blunted phasic autonomic response to errors indexes age-related deficits in error awareness. Neurobiol. Aging 71, 13–20. doi:10.1016/j.neurobiolaging.2018.06.019.

Yeung, N., Botvinick, M.M., Cohen, J.D., 2004. The neural basis of error detection: conflict monitoring and the error-related negativity. Psychol. Rev. 111, 931–959. doi:10.1037/0033-295X.111.4.931.

Yeung, N., Cohen, J.D., 2006. The Impact of Cognitive Deficits on Conflict MonitoringPredictable Dissociations Between the Error-Related Negativity and N2. Psychol. Sci. 17, 164–171.

Yeung, N., Summerfield, C., 2014. Shared mechanisms for confidence judgements and error detection in human decision making. In: The Cognitive Neuroscience of Metacognition. Springer-Verlag, Berlin Heidelberg, pp. 147–167. doi:10.1007/978-3-642-45190-4_7.

3.1.2. Supplement

Neural correlates of metacognition across the adult lifespan: Supplementary Material

Helen Overhoff[a,b,c*], Yiu Hong Ko[a,b,c], Daniel Feuerriegel[b], Gereon R. Fink[a,d], Jutta Stahl[c],
Peter H. Weiss[a,d], Stefan Bode[b], & Eva Niessen[c]

[a]Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Research
Centre Jülich, Leo-Brandt-Str. 5, 52425 Jülich, Germany

[b]Melbourne School of Psychological Sciences, University of Melbourne, Parkville Campus,
Parkville 3010, Victoria, Australia

[c]Department of Individual Differences and Psychological Assessment, University of
Cologne, Pohligstr. 1, 50969 Cologne, Germany

[d]Department of Neurology, Faculty of Medicine, University Hospital Cologne, University
of Cologne, Kerpener Str. 62, 50937 Cologne, Germany

*Corresponding author. Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Research Centre Jülich, Leo-Brandt-Str. 5, 52425 Juelich, Germany.

Email: h.overhoff@fz-juelich.de

Supplementary Material List of Contents

S1. Mixed effects regression model structures and coefficients for behavioural analyses

For the analysis of behavioural parameters, data were analysed using linear and generalised linear mixed effects models. We always used the between-subjects factor age as the regressor. The within-subject factor of interest was either accuracy (error, correct) or (pooled) confidence (3 levels). We fitted random intercepts for participants and, if possible, random slopes by participant for the within-subject factor of interest. For linear mixed models, *F* statistics are reported, and degrees of freedom were estimated by Satterthwaite's approximation, and for generalised linear mixed models, X^2 statistics are reported.

Significant effects of accuracy or confidence were followed up by pairwise comparisons between error and correct trials or across confidence levels using paired-samples *t*-tests for linear mixed models and *Z*-tests for generalised linear models. Significant interactions were followed up by (generalised) linear mixed regressions, separately for each level of a given within-subject factor to assess potential effects of age. These follow-up tests were chosen because our main interest was in the differential relations between accuracy, confidence, and behaviour across the lifespan rather than between the levels. Post-hoc test results were compared against Holm-corrected alpha levels to account for multiple comparisons.

Error rate (ER)

For the analysis of the error rate, we fitted a generalised linear mixed effects model (binomial family, logit function) testing for effects of confidence and age on accuracy. The variable of age was centred and scaled.

```
Accuracy ~ confidence * age + (1 | sbj)
```

Analysis of Deviance Table with Wald tests.

Predictor	*df*	X^2	*p*
Confidence	2	2200.020	**<.001**
Age	1	4.704	**.030**
Confidence*Age	2	168.125	**<.001**

Regression coefficients for the predictor of age

	Estimate	Std. Error	*z* Ratio
Age	0.37	0.169	2.169

Post-hoc test of contrasts between confidence levels.

Contrast	Estimate	Std. Error	*z* Ratio	*p*
Low – medium	-3.88	0.071	28.872	**<.001**
Low – high	-5.93	0.182	-10.103	**<.001**
Medium – high	-2.05	0.147	-26.395	**<.001**

```
Accuracy [low conf] ~ age + (1 | sbj)

Accuracy [medium conf] ~ age + (1 | sbj)

Accuracy [high conf] ~ age + (1 | sbj)
```

Post-hoc Analysis of Deviance Table with Wald tests and regression coefficients for the predictor of age for each level of confidence.

Predictor	*df*	X^2	Estimate	Std. Error	*p*
Age [low conf]	1	0.898	0.01	0.012	.343
Age [medium conf]	1	3.467	-0.01	0.007	.063
Age [high conf]	1	37.664	-0.05	0.009	**<.001**

Response time (RT)

For the analysis of RTs, we fitted linear mixed effects models testing for effects of accuracy and age or confidence and age on RTs, respectively.

```
RT ~ accuracy * age + (accuracy | sbj)
```

Analysis of Variance Table with Satterthwaite's method.

Predictor	*Num df*	*Den df*	*F*	*p*
Accuracy	1	61.6	5.572	**.021**
Age	1	62.9	17.358	**<.001**
Accuracy*Age	1	56.5	2.846	.097

Regression coefficients for the predictor of age

	Estimate	Std. Error	*t* Ratio
Age	2.22	0.734	3.023

Post-hoc test of contrasts between error and correct.

Contrast	Estimate	Std. Error	*t* Ratio	*p*
Error – correct	17.9	7.15	2.498	**.015**

```
RT ~ confidence * age + (confidence | sbj)
```

Analysis of Variance Table with Satterthwaite's method.

Predictor	*Num df*	*Den df*	*F*	*p*
Confidence	2	61.8	27.291	**<.001**
Age	1	63.7	13.305	**<.001**
Confidence*Age	2	56.6	5.187	**.009**

Regression coefficients for the predictor of age

	Estimate	Std. Error	*t* Ratio
Age	3.07	0.845	3.637

Post-hoc test of contrasts between confidence levels.

Contrast	Estimate	Std. Error	*t* Ratio	*p*
Low – medium	-72.3	10.40	-6.952	**<.001**
Low – high	25.3	9.48	2.673	**.010**
Medium – high	97.6	6.74	14.496	**<.001**

```
RT [low conf] ~ age + (1 | sbj)

RT [medium conf] ~ age + (1 | sbj)

RT [high conf] ~ age + (1 | sbj)
```

Post-hoc Analysis of Deviance Table with Wald tests and regression coefficients for the predictor of age for each level of confidence.

Predictor	*Num df*	*Den df*	*F*	Estimate	Std. Error	*p*
Age [low conf]	1	68.4	13.592	3.19	0.866	**<.001**
Age [medium conf]	1	53.9	3.634	1.38	0.725	.062
Age [high conf]	1	62.4	18.358	2.47	0.578	**<.001**

Confidence

For the analysis of confidence, we fitted a linear mixed effects model testing for the effects of accuracy and age on confidence ratings.

```
Confidence ~ accuracy * age + (accuracy | sbj)
```

Analysis of Variance Table with Satterthwaite's method.

Predictor	*Num df*	*Den df*	*F*	*p*
Accuracy	1	63.4	162.928	**<.001**
Age	1	62.9	2.078	.154
Accuracy*Age	1	62.4	37.361	**<.001**

Post-hoc test of contrasts between error and correct.

Contrast	Estimate	Std. Error	*t* Ratio	*p*
Error – correct	-0.986	0.046	-21.252	**<.001**

```
Confidence [error] ~ age + (1 | sbj)
```

```
Confidence [correct] ~ age + (1 | sbj)
```

Post-hoc Analysis of Deviance Table with Wald tests and regression coefficients for the predictor of age for error and correct trials.

Predictor	*Num df*	*Den df*	*F*	Estimate	Std. Error	*p*
Age [error]	1	68.4	17.977	0.01	0.003	**<.001**
Age [correct]	1	53.9	23.816	-0.01	0.001	**<.001**

Behavioural adaptation

For the analysis of behavioural adjustments, we fitted linear mixed effects models testing for effects of accuracy and age or confidence and age on response caution, respectively.

```
Response caution ~ accuracy * age + (accuracy | sbj)
```

Analysis of Variance Table with Satterthwaite's method.

Predictor	*Num df*	*Den df*	*F*	*p*
Accuracy	1	55.9	12.366	**<.001**
Age	1	62.4	2.141	.148
Accuracy*Age	1	43.9	6.709	**.013**

Post-hoc test of contrasts between error and correct.

Contrast	Estimate	Std. Error	*t* Ratio	*p*
Error – correct	27.6	7.26	3.808	**<.001**

```
Response caution [error] ~ age + (1 | sbj)
```

```
Response caution [correct] ~ age + (1 | sbj)
```

Post-hoc Analysis of Deviance Table with Wald tests and regression coefficients for the predictor of age for error and correct trials.

Predictor	*Num df*	*Den df*	*F*	Estimate	Std. Error	*p*
Age [error]	1	60.5	3.836	-1.82	45.978	.055
Age [correct]	1	61.9	0.584	-0.60	0.782	.448

```
Response caution ~ confidence * age + (confidence | sbj)
```

Analysis of Variance Table with Satterthwaite's method.

Predictor	*Num df*	*Den df*	*F*	*p*
Confidence	2	54.9	7.306	**.002**
Age	1	63.5	2.935	.092
Confidence*Age	2	50.4	2.837	.068

Post-hoc test of contrasts between confidence levels.

Contrast	Estimate	Std. Error	*t* Ratio	*p*
Low – medium	13.0	13.42	0.969	.336
Low – high	42.8	10.67	4.013	**<.001**
Medium – high	29.8	8.78	3.399	**.002**

S2. Mixed effects regression model structures and coefficients for electrophysiological analyses

For the analysis of electrophysiological parameters, data were analysed fitting the same models and performing the same post-hoc tests as for the behavioural data

$N_{e/c}$ amplitudes

For the analysis of the $N_{e/c}$, we fitted a linear mixed effects model testing for effects of accuracy and age on the mean amplitude including all trials.

```
Ne/c ~ accuracy * age + (accuracy | sbj)
```

Analysis of Variance Table with Satterthwaite's method.

Predictor	*Num df*	*Den df*	*F*	*p*
Accuracy	1	38.6	9.054	**.005**
Age	1	31.3	3.484	.067
Accuracy*Age	1	31.8	5.472	**.026**

Post-hoc test of contrasts between error and correct.

Contrast	Estimate	Std. Error	*t* Ratio	*p*
Error – correct	-0.020	0.007	-2.842	**.007**

```
Ne [error] ~ age + (1 | sbj)

Nc [correct] ~ age + (1 | sbj)
```

Post-hoc Analysis of Deviance Table with Wald tests and regression coefficients for the predictor of age for error and correct trials.

Predictor	*Num df*	*Den df*	*F*	Estimate	Std. Error	*p*
Age [error]	1	55.4	5.030	0.00	0.001	**.029**
Age [correct]	1	62.6	1.124	0.00	0.001	.293

For the analysis of the N_e of errors, we fitted a linear mixed effects model testing for effects of confidence and age on the mean amplitude including only errors.

```
Ne ~ confidence * age + (1 | sbj)
```

Analysis of Variance Table with Satterthwaite's method.

Predictor	*Num df*	*Den df*	*F*	*p*
Confidence	2	2706.4	4.007	**.018**
Age	1	57.4	4.068	**.048**
Confidence*Age	2	2731.5	3.662	**.026**

Regression coefficients for the predictor of age

	Estimate	Std. Error	*t* Ratio
Age	0.00	0.000	3.000

Post-hoc test of contrasts between confidence levels.

Contrast	Estimate	Std. Error	*t* Ratio	*p*
Low – medium	-0.010	0.013	1.696	.180
Low – high	42.8	10.67	2.795	**.016**
Medium – high	29.8	8.78	0.908	.364

```
Ne [low conf] ~ age + (1 | sbj)
```

```
Ne [medium conf] ~ age + (1 | sbj)
```

```
Ne [high conf] ~ age + (1 | sbj)
```

Post-hoc Analysis of Deviance Table with Wald tests and regression coefficients for the predictor of age for each level of confidence.

Predictor	*Num df*	*Den df*	*F*	Estimate	Std. Error	*p*
Age [low conf]	1	58.1	9.735	0.00	0.001	**.003**
Age [medium conf]	1	37.0	2.394	0.00	0.039	.130
Age [high conf]	1	43.9	0.517	0.00	0.001	.476

For the analysis of the N_c of correct responses, we fitted a linear mixed effects model testing for effects of confidence and age on the mean amplitude including only correct trials.

```
Nc ~ confidence * age + (1 | sbj)
```

Analysis of Variance Table with Satterthwaite's method.

Predictor	*Num df*	*Den df*	*F*	*p*
Confidence	2	16678.6	0.428	.652
Age	1	220.9	2.573	.110
Confidence*Age	2	16565.6	1.145	.318

$P_{e/c}$ amplitudes

For the analysis of the $P_{e/c}$, we fitted a linear mixed effects model testing for effects of accuracy and age on the mean amplitude including all trials.

```
Pe/c ~ accuracy * age + (accuracy | sbj)
```

Analysis of Variance Table with Satterthwaite's method.

Predictor	*Num df*	*Den df*	*F*	*p*
Accuracy	1	55.3	10.378	**.002**
Age	1	62.5	0.025	.876
Accuracy*Age	1	49.2	6.443	**.014**

Post-hoc test of contrasts between error and correct.

Contrast	Estimate	Std. Error	*t* Ratio	*p*
Error – correct	0.031	0.011	2.799	**.007**

```
Pe [error] ~ age + (1 | sbj)
```

```
Pc [correct] ~ age + (1 | sbj)
```

Post-hoc Analysis of Deviance Table with Wald tests and regression coefficients for the predictor of age for error and correct trials.

Predictor	*Num df*	*Den df*	*F*	Estimate	Std. Error	*p*
Age [error]	1	58.5	0.976	-0.00	0.001	.328
Age [correct]	1	63.5	1.562	0.00	0.000	.219

For the analysis of the P_e of errors, we fitted a linear mixed effects model testing for effects of confidence and age on the mean amplitude including only errors.

```
Pe ~ confidence * age + (confidence | sbj)
```

Analysis of Variance Table with Satterthwaite's method.

Predictor	*Num df*	*Den df*	*F*	*p*
Confidence	2	57.5	0.810	.450
Age	1	59.2	0.425	.517
Confidence*Age	2	40.6	0.294	.747

For the analysis of the P_c of correct responses, we fitted a linear mixed effects model testing for effects of confidence and age on the mean amplitude including only correct trials.

```
Pc ~ confidence * age + (1 | sbj)
```

Analysis of Variance Table with Satterthwaite's method.

Predictor	*Num df*	*Den df*	*F*	*p*
Confidence	2	16709	0.313	.732
Age	1	181	1.740	.189
Confidence*Age	2	16650	1.364	.256

S3. Modulation of ERPs by confidence, independent of accuracy

In our main analysis, we fitted linear mixed effects models to the $N_{e/c}$ and $P_{e/c}$ amplitudes of all trials with the within-subject factor accuracy and the between-subject factor age. The amplitudes of both ERPs were larger for errors than correct responses, and the $N_{e/c}$ amplitude decreased with age for errors. As both components have further been shown to be sensitive to variations in confidence (Boldt & Yeung, 2015), we additionally computed the $N_{e/c}$ and $P_{e/c}$ amplitudes in relation to reported confidence for errors and correct trials combined (three levels: ‘surely wrong’, ‘unsure’, ‘surely correct’). Here, we provide the results for the linear mixed effects regression analyses including confidence instead of accuracy as the within-subject factor.

$N_{e/c}$ amplitude

For the analysis of the $N_{e/c}$, we fitted a linear mixed effects model testing for effects of confidence and age on the mean amplitude including all trials.

```
Ne/c ~ confidence * age + (confidence | sbj)
```

Analysis of Variance Table with Satterthwaite’s method.

Predictor	*Num df*	*Den df*	*F*	*p*
Confidence	2	47.2	7.057	**.002**
Age	1	60.6	4.703	**.034**
Confidence*Age	2	40.4	4.989	**.012**

Regression coefficients for the predictor of age

	Estimate	Std. Error	*t* Ratio
Age	0.00	0.001	2.960

Post-hoc test of contrasts between confidence levels.

Contrast	Estimate	Std. Error	*t* Ratio	*p*
Low – medium	-0.030	0.011	-2.811	**.022**
Low – high	-0.026	0.011	-2.449	**.038**
Medium – high	-0.004	0.007	0.649	.519

```
Ne/c [low conf] ~ age + (1 | sbj)
```

```
Ne/c [medium conf] ~ age + (1 | sbj)
```

```
Ne/c [high conf] ~ age + (1 | sbj)
```

Post-hoc Analysis of Deviance Table with Wald tests and regression coefficients for the predictor of age for each level of confidence.

Predictor	*Num df*	*Den df*	*F*	Estimate	Std. Error	*p*
Age [low conf]	1	58.1	11.353	0.00	0.001	**.001**
Age [medium conf]	1	45.2	0.453	0.00	0.001	.505
Age [high conf]	1	64.1	0.355	0.00	0.001	.553

$P_{e/c}$ amplitude

For the analysis of the $P_{e/c}$, we fitted a linear mixed effects model testing for effects of confidence and age on the mean amplitude including all trials.

```
Pe/c ~ confidence * age + (confidence | sbj)
```

Analysis of Variance Table with Satterthwaite's method.

Predictor	*Num df*	*Den df*	*F*	*p*
Confidence	2	52.1	3.817	**.028**
Age	1	63.1	0.076	.783
Confidence*Age	2	47.4	1.452	.244

Post-hoc test of contrasts between confidence levels.

Contrast	Estimate	Std. Error	*t* Ratio	*p*
Low – medium	0.037	0.018	2.091	.083
Low – high	0.053	0.015	3.581	**.002**
Medium – high	0.016	0.010	1.660	.103

Together, these results replicate previous findings of a confidence-related modulation of the $N_{e/c}$ and $P_{e/c}$ amplitudes (Boldt & Yeung, 2015; Scheffers & Coles, 2000). Moreover, we provide evidence, for the first time, that the $N_{e/c}$, but not the $P_{e/c}$ was differently modulated by ageing across confidence levels. Notably, it has to be considered that the percentage of errors within each confidence level varied substantially between participants and across the lifespan. However, the same holds for the opposite conclusion, that is, a potential modulation of the ERP amplitudes by accuracy is always inherently connected to confidence (e.g., Fleming et al., 2012). Therefore, the more robust analysis, in our opinion, is the separate examination of correct and incorrect trials, which we report in the main article.

S4. Stimulus-related ERPs of conflict processing

When investigating age-related alterations in neural correlates of response evaluation, the interval between stimulus presentation and response is also informative – in particular the N2 and the P300 components of the ERP. These have been related to error processing as indexing stimulus conflict monitoring (N2) and error-related attention reallocation (P300; Groom & Cragg, 2015; Polich, 2007; Yeung & Cohen, 2006). Research has shown a decline of both components in older age (Korsch et al., 2016; Lucci et al., 2013). Assessing the modulation of these components by age allowed us to draw conclusions about the specificity of potential modulations of the $N_{e/c}$ and $P_{e/c}$ in our metacognitive task. We, therefore, additionally computed the N2 and the P300 components using the stimulus-locked data.

The epochs were cut at 1,500 ms after target stimulus presentation, and the preprocessing was equivalent to the response-locked data. The N2 was quantified as the mean amplitude around the negative peak latency (± 50 ms) of the grand-average ERP in the time window from 150 to 300 ms at Cz, and the P300 around the positive peak latency (± 50 ms) of the grand-average ERP in the time window from 200 to 500 ms at POz (Groom & Cragg, 2015; Klawohn et al., 2020; Polich, 2007). The latencies were retrieved for errors and correct responses, respectively.

N2 amplitudes

For the analysis of the N2, we fitted a linear mixed effects model testing for effects of accuracy and age on the mean amplitude including all trials.

```
N2 ~ accuracy * age + (accuracy | sbj)
```

Analysis of Variance Table with Satterthwaite's method.

Predictor	*Num df*	*Den df*	*F*	*p*
Accuracy	1	2188.4	0.001	.974
Age	1	58.3	1.586	.220
Accuracy*Age	1	1487.4	0.005	.942

P300 amplitudes

For the analysis of the P300, we fitted a linear mixed effects model testing for effects of accuracy and age on the mean amplitude including all trials.

```
P300 ~ accuracy * age + (accuracy | sbj)
```

Analysis of Variance Table with Satterthwaite's method.

Predictor	*Num df*	*Den df*	*F*	*p*
Accuracy	1	89.6	0.012	.914
Age	1	60.8	0.281	.598
Accuracy*Age	1	67.0	0.184	.670

The results suggests that both the monitoring of stimulus conflict and the attention-related evaluation of conflict were comparable across the lifespan. This means that age-related differences in early stimulus-related conflict monitoring do not account for subsequent modulations of response processing.

S5. Electrophysiological analyses using untransformed ERP data

In addition to the analysis of the CSD-transformed ERP data reported in the manuscript, we computed the same analyses for the analysis of accuracy using untransformed raw data.

$N_{e/c}$ amplitudes

For the analysis of the $N_{e/c}$, we fitted a linear mixed effects model testing for effects of accuracy and age on the mean amplitude including all trials.

$N_{e/c}$ ~ accuracy * age + (accuracy | sbj)

Analysis of Variance Table with Satterthwaite's method.

Predictor	*Num df*	*Den df*	*F*	*p*
Accuracy	1	48.4	23.326	**<.001**
Age	1	59.3	0.297	.588
Accuracy*Age	1	41.4	10.754	**.002**

Post-hoc test of contrasts between error and correct.

Contrast	Estimate	Std. Error	*t* Ratio	*p*
Error – correct	-1.46	0.257	-5.698	**<.001**

N_{e} [error] ~ age + (1 | sbj)

N_{c} [correct] ~ age + (1 | sbj)

Post-hoc Analysis of Deviance Table with Wald tests for error and correct trials.

Predictor	*Num df*	*Den df*	*F*	*p*
Age [error]	1	51.9	2.728	.105
Age [correct]	1	62.6	1.424	.237

$P_{e/c}$ amplitudes

For the analysis of the $N_{e/c}$, we fitted a linear mixed effects model testing for effects of accuracy and age on the mean amplitude including all trials.

$P_{e/c}$ ~ accuracy * age + (accuracy | sbj)

Analysis of Variance Table with Satterthwaite's method.

Predictor	*Num df*	*Den df*	*F*	*p*
Accuracy	1	60.6	5.338	**.024**
Age	1	61.4	0.595	.443
Accuracy*Age	1	54.2	3.143	.082

Post-hoc test of contrasts between error and correct.

Contrast	Estimate	Std. Error	*t* Ratio	*p*
Error – correct	0.746	0.351	2.124	**.038**

In sum, the pattern of results for the raw ERP data was very similar to the results for the CSD-transformed data. The only discrepancy was that two effects did not become significant using the raw data, namely the effect of age in the post-hoc analysis for the N_e of errors, and the interaction between age and accuracy for the $P_{e/c}$ amplitude. As this effect was not large in our reported analysis either, the increased noise in the raw data might have concealed the effect in this analysis.

S6. Bayesian statistics for reported null-effects

In our manuscript, we are reporting frequentist statistics for all analyses in order to keep the statistical framework consistent. However, we computed additional analyses reporting Bayes factors for all null findings, because they constitute an essential part of our conclusion (i.e., null effects of age for response caution and $P_{e/c}$ amplitude).

We ran Bayesian statistical analyses using the package BayesFactor in R (version 0.9.12-4.2; Morey and Rouder, 2018) to assess the extent to which our data support the null effects. In order to examine the null effects of age for response caution and $P_{e/c}$ amplitude, we compared the full models including the within-subject factor of interest (accuracy or confidence) and the between-subject factor age to a null model including only the within-subject factor.

For response caution, we tested the hypothesis that response caution is modulated by accuracy, age, and their interaction against the null hypothesis that it was only modulated by accuracy. We found anecdotal evidence in favour of the null hypothesis (BF_{01} = 2.055). Comparing the model predicting response caution by the factors confidence and age to the model including only age, we found strong evidence supporting the null hypothesis (BF_{01} = 238.612). This suggests that the modulation of response caution was indeed similar across the lifespan.

For the $P_{e/c}$, we assessed evidence for a modulation by age of all trials combined. Here, the model including the interaction term of accuracy and age was around five times more likely given the data than the null model (BF_{10} = 5.264). This is mirroring the significant interaction we found in the analysis reported in the manuscript. For the modulation of errors by confidence and age, we found strong evidence against an effect of age on the Pe amplitude (BF_{01} = 614.830). The same was true for the age effect on the Pc amplitude of correct responses (BF_{01} = 1294.515).

Taken together, these results suggest that our data robustly support the null effects of age on the confidence modulation of response caution and the $P_{e/c}$ amplitude.

S7 Supplementary material reference list

Boldt, A., & Yeung, N. (2015). Shared neural markers of decision confidence and error detection. *The Journal of Neuroscience, 35*(8), 3478–3484. https://doi.org/10.1523/JNEUROSCI.0797-14.2015

Fleming, S. M., Dolan, R. J., & Frith, C. D. (2012). Metacognition: Computation, biology and function. Philosophical *Transactions of the Royal Society B: Biological Sciences, 367*(1594), 1280–1286. https://doi.org/10.1098/rstb.2012.0021

Groom, M. J., & Cragg, L. (2015). Differential modulation of the N2 and P3 event-related potentials by response conflict and inhibition. *Brain and Cognition, 97*, 1–9. https://doi.org/10.1016/j.bandc.2015.04.004

Klawohn, J., Santopetro, N. J., Meyer, A., & Hajcak, G. (2020). Reduced P300 in depression: Evidence from a flanker task and impact on ERN, CRN, and Pe. *Psychophysiology, 57*(4), 1–11. https://doi.org/10.1111/psyp.13520

Korsch, M., Frühholz, S., & Herrmann, M. (2016). Conflict-specific aging effects mainly manifest in early information processing stages-an ERP study with different conflict types. *Frontiers in Aging Neuroscience, 8*, 1–12. https://doi.org/10.3389/fnagi.2016.00053

Lucci, G., Berchicci, M., Spinelli, D., Taddei, F., & Di Russo, F. (2013). The Effects of Aging on Conflict Detection. *PLoS ONE, 8*(2). https://doi.org/10.1371/journal.pone.0056566

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes Factors for Common Designs.* (R package version 0.9.12-4.2). https://cran.r-project.org/package=BayesFactor

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Scheffers, M. K., & Coles, M. G. H. (2000). Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance, 26*(1), 141–151. https://doi.org/10.1037/0096-1523.26.1.141

Yeung, N., & Cohen, J. D. (2006). The impact of cognitive deficits on conflict monitoring predictable dissociations between the error-related negativity and N2. *Psychological Science, 17*(2), 164–171. https://doi.org/10.1111/j.1467-9280.2006.01680.x

3.1.3. Summary

In this paper, we addressed the research questions 1 to 3 of this thesis. To summarise, we found that metacognitive accuracy, quantified as Phi, decreased gradually across the lifespan beyond the observed decline in primary task performance, and was characterised by more conservative confidence ratings in older adults. Despite this strong decline, participants adjusted their response policies similarly across age: after errors and after trials that were rated with low confidence, they responded slower and more accurately. Regarding the electrophysiological activity, we showed larger $N_{e/c}$ and $P_{e/c}$ amplitudes for lower confidence ratings for errors, but not for correct responses. Surprisingly, the P_e amplitude did not show the expected age-related decline for errors that were likely detected (i.e., rated with low confidence). However, the N_e amplitude of such trials was significantly smaller with older age and thus, less sensitive to different levels of confidence.

In the context of this thesis, we can conclude that the age-related decline in metacognitive performance that was observed in a signal detection task (Palmer et al., 2014) extends to a complex version of a conflict task typically used in studies of error monitoring. Interestingly, older adults were more uncertain regarding their performance than younger adults. An age-related decline was observed in a neural correlate of early error monitoring, while no effect of age was found for the modulation of the P_e by confidence and the adjustment of subsequent behaviour.

3.2. Study 2

3.2.1. Aims

This manuscript presents the equivalent to Study 2 of this thesis. In order to shed light on the mechanisms underpinning metacognitive processing, we explored how age impacted on the behavioural correlates of confidence. The study is based on the same data as Study 1. Here, we specifically focussed on the recorded response time (RT) and peak force (PF) of the first-order decisions and modelled how these parameters were related to confidence judgements across correct and incorrect responses and across age. While results from previous studies suggest that information about the action of reporting a decision has a direct effect on confidence judgements, independent of objective accuracy (e.g., Siedlecka et al., 2021; Turner, Angdias, et al., 2021), we were primarily interested in the relationship between metacognition and naturally occurring response dynamics in our novel paradigm. Observing a marked effect of age on metacognitive performance in Study 1, we expected altered patterns of these relationships with older age.

This manuscript has been submitted to *Frontiers in Aging Neuroscience* and is presented in the submitted version, formatted according to APA guidelines (7th edition).

3.2.2. Manuscript

The relationship between response dynamics and the formation of confidence varies across the lifespan

Helen Overhoff[a,b,c*], Yiu Hong Ko[a,b,c], Gereon R. Fink[a,d], Jutta Stahl[c], Peter H. Weiss[a,d], Stefan Bode[b], & Eva Niessen[c]

[a]Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Research Centre Jülich, Leo-Brandt-Str. 5, 52425 Jülich, Germany

[b]Melbourne School of Psychological Sciences, University of Melbourne, Parkville Campus, Parkville 3010, Victoria, Australia

[c]Department of Individual Differences and Psychological Assessment, University of Cologne, Pohligstr. 1, 50969 Cologne, Germany

[d]Department of Neurology, Faculty of Medicine, University Hospital Cologne, University of Cologne, Kerpener Str. 62, 50937 Cologne, Germany

*Corresponding author. Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Research Centre Jülich, Leo-Brandt-Str. 5, 52425 Juelich, Germany.

Email: h.overhoff@fz-juelich.de

Abstract

Accurate metacognitive judgements, such as forming a confidence judgement, are crucial for goal-directed behaviour but decline with older age. Besides changes in the sensory processing of stimulus features, there might also be changes in the motoric aspects of giving responses that account for age-related changes in confidence. In order to assess the association between confidence and response parameters across the adult lifespan, we measured response times and peak forces in a four-choice flanker task with subsequent confidence judgements. In 65 healthy adults from 20 to 76 years of age, we showed divergent associations of each measure with confidence, depending on decision accuracy. Participants indicated higher confidence after faster responses in correct but not incorrect trials. They also indicated higher confidence after less forceful responses in errors but not in correct trials. Notably, these associations were age-dependent as the relationship between confidence and response time was more pronounced in older participants, while the relationship between confidence and response force decayed with age. Our results add to the notion that confidence is related to response parameters and demonstrate noteworthy changes in the observed associations across the adult lifespan. These changes potentially constitute an expression of general age-related deficits in performance monitoring or, alternatively, index a failing mechanism in the computation of confidence in older adults.

Keywords: ageing, confidence, metacognitive accuracy, response parameters, response force

1. Introduction

Humans can report a subjective sense of confidence that is closely related to the accuracy of their actions. This ongoing monitoring of decisions and their execution is called metacognition and includes the evaluation of behaviour and the detection of occurring errors (Fleming & Dolan, 2012). Accurate metacognitive judgements should lead to adaptive behaviour adjustments and are thus crucial for all activities. Undetected errors (i.e., incorrect metacognitive judgements) might have severe implications for real-life scenarios because they may not trigger the required adjustments for future actions and decisions (Wessel et al., 2018).

1.1. Age-related decline in metacognitive accuracy

Research on metacognitive performance across the adult lifespan has consistently pointed towards a decline in older age. When participants were asked to report committed errors in an easy choice-reaction task, the detection rates declined with age, even when task performance was comparable (Harty et al., 2017; Niessen et al., 2017). In our previous publication, using the same dataset described in this study (Overhoff et al., 2021), we asked participants to rate their confidence after each decision on a four-point scale, and we concordantly revealed a decline in metacognitive performance across the lifespan. The accuracy of these ratings decreased gradually with higher age, reflecting that older adults were less aware of their errors and rated correct responses with lower confidence compared to younger adults (see also Palmer et al., 2014). As of today, the question of which factors are related to this selective decline remains open.

1.2. Computation of confidence

In order to understand the age-related decline in metacognitive performance, it is essential to understand the basic mechanisms underlying the computation of confidence. It is still unclear which information is used to compute confidence, and how input from different sources is weighted (Charles & Yeung, 2019; Feuerriegel et al., 2021). For instance, confidence has been related to the strength of stimulus evidence, stimulus discriminability (Charles & Yeung, 2019; Turner, Feuerriegel, et al., 2021; Yeung & Summerfield, 2012), or instructed time pressure (Vickers & Packer, 1982). Furthermore, growing evidence suggests that the interoceptive feedback of a motor action while giving a response might be another source of information contributing to the formation of confidence about the decision (Fleming et al., 2015; Gajdos et al., 2019; Kiani et al., 2014; Palser et al., 2018; Siedlecka et al., 2021; Turner, Angdias, et al., 2021). Fleming and colleagues (2015) investigated the interaction of confidence and motor-related activity by delivering single-pulse transcranial magnetic stimulation (TMS) to the dorsal premotor cortex. This perturbation did not affect task performance, but crucially, it did affect the accuracy of subsequent confidence judgements, i.e., the degree to which the judgements matched the observed performance. This finding indicates that action-specific cortical activations might contribute to confidence. In line with this assumption, confidence ratings have been shown to be more accurate if the preceding decision required a motor action (Pereira et al., 2020; Siedlecka et al., 2021). For instance, Siedlecka and colleagues (2021) recently showed that metacognitive accuracy was higher after decisions requiring a key press than decisions which were indicated without a motor action. Taken together, these findings suggest that features of the motor response indicating a given decision might influence the confidence ratings about this decision. Therefore, further investigations of how confidence is reflected in different response parameters are warranted.

1.3. Differential relationship between confidence and response parameters

A response can be characterised by different dimensions. The most commonly used output variable is time, usually response time or movement time. A robust finding across studies is a negative relationship between response times for the initial decision and subsequent confidence ratings (Fleming et al., 2010; Kiani et al., 2014; Rahnev et al., 2020). Intuitively, one might assume that the degree of confidence is expressed in the time taken to make the decision, i.e., the less confident we are about a decision, the longer it should take to respond. However, another possible explanation is that the monitoring system uses the interoceptive signal of a movement produced by the response as an informative cue about the difficulty of the decision (Fleming & Daw, 2017; Kiani et al., 2014). Accordingly, if an easy decision led to a fast response, the internal read-out could boost subjective confidence. A recent study provided evidence for the directional effect of movement time (i.e., the time from lifting to dropping a marble) on confidence (Palser et al., 2018). In this study, movement speed was experimentally manipulated by instructing participants to move faster than they naturally would, and this manipulation resulted in declined metacognitive accuracy.

Nevertheless, temporal parameters do not capture all aspects of a movement. For instance, subthreshold motor activity (i.e., partial responses) cannot be detected by classical RT recordings but rather by recording muscle activity. However, partial responses have also been shown to affect reported confidence (Ficarella et al., 2019; Gajdos et al., 2019). An informative motor parameter of a response is the applied force, which is often measured in its peak force, i.e., the maximum exerted force during a response action. Notably, peak force and response time index distinct processes as they show divergent behavioural patterns (i.e., small to no correlation) across experimental manipulations (Cohen & van Gaal, 2014; Franz & Miller, 2002; Stahl & Rammsayer, 2005).

Contrary to the well-known negative relationship between confidence and response time, the association between confidence and response force has rarely been investigated. Recently, Turner and colleagues (2021a) examined this relationship by explicitly manipulating the degree of physical effort that had to be exerted to give a response. When participants were prompted to submit their response to a perceptual decision with varying force levels, participants reported higher confidence in their decisions when their response peak force was higher. Notably, requiring participants to produce a specific (and comparably high) degree of force (as mandated in the experiment) is fundamentally different from measuring naturally occurring force patterns of a response (in terms of a dependent measure). The latter was done, for example, in a study by Bode and Stahl (2014), who found that naturally occurring peak force was lower in errors compared to correct responses. It was suggested that this might indicate a process in which low force in error trials signifies an unsuccessful attempt to stop the already initiated response, which requires early and fast error detection (Bode & Stahl, 2014; Y. T. Ko et al., 2012; Stahl et al., 2020). However, error detection or confidence was not directly assessed, rendering comparison between these two studies difficult.

The relationship between different response parameters and confidence has not been systematically assessed in the context of healthy ageing. While response and movement times are slower and more variable with older age, findings of age-related changes in response force are inconsistent (Bunce et al., 2004; Dully et al., 2018; Salthouse, 2000). Some studies showed delayed and altered electrophysiological signatures of motor processing in older age (e.g., lateralised readiness potential (LRP)/ movement-related potential (MRP) and mu/ beta desynchronisation; Falkenstein et al., 2006; Quandt et al., 2016; Sailer et al., 2000). In contrast, electromyographic or force recordings of motor responses revealed similar patterns in younger and older adults (Dully et al., 2018; Falkenstein et al., 2006; Van Der Lubbe et al., 2002; Yordanova et al., 2004). Notably, these studies did not assess error awareness or confidence. Therefore, it is warranted to specifically examine the associations between confidence and response time and between confidence and response force and to investigate whether these associations change across the adult lifespan.

1.4. Objectives

The present study constitutes the first comprehensive assessment of the association between metacognitive accuracy and two main response parameters across the adult lifespan. We intended to answer the following questions: first, what are the relationships between decision confidence and response time on the one hand, and peak force of a response (as it naturally occurs, i.e., without specific instruction or experimental manipulation) on the other hand? Second, do these relationships between confidence and response parameters change with age? Additionally, we were interested in investigating the potential moderating effect of accuracy because many studies on decision confidence only assessed the relationship between a given response parameter and confidence in correct responses. However, we can only understand the computation of confidence when considering errors (Charles & Yeung, 2019; Dotan et al., 2018; Peters et al., 2017). Concerning response time, for instance, the well-known negative relationship with confidence is inverted for errors when the confidence rating is allowed to indicate error detection (i.e., a rating scale was used that ranged from certainty in being correct to certainty in being wrong; Pereira et al., 2020).

We expected significant associations between confidence judgements and parameters of the response (Fleming et al., 2015; Pereira et al., 2020; Rahnev et al., 2020; Turner, Feuerriegel, et al., 2021). In particular, response time was expected to decrease with higher confidence for correct trials (Dotan et al., 2018; Kiani et al., 2014; Rahnev et al., 2020) and to increase with higher confidence for errors (Pereira et al., 2020). We tentatively hypothesised a positive relationship between response force and confidence for errors and correct responses (Bode & Stahl, 2014; Y. T. Ko et al., 2012; Turner, Angdias, et al., 2021).

Most importantly, we intended to explore age-related changes in the associations between response parameters and confidence without having a priori hypotheses about the direction of possible effects due to a lack of previous studies on this topic. If we find divergent patterns across the lifespan, this might encourage research on the causal relationship between response parameters and confidence.

2. Methods

2.1. Participants

Eighty-two participants were recruited and received monetary compensation for their participation in the experiment. Data from seventeen participants had to be discarded due to: symptoms of depression (N = 1, Beck's Depression Inventory score higher than 17; BDI; Hautzinger, 1991), poor behavioural performance (N = 8, more than 30% invalid trials, error rate higher than chance, here 25%), or a behavioural pattern that was indicative of an insufficient understanding or implementation of task demands (N = 8, inspection of individual datasets for a combination of errors in the colour discrimination test described below, near chance task performance, frequent invalid trials, and biased use of single response keys). This resulted in a final sample for analysis of sixty-five healthy, right-handed adults (age = 45.5 ± 2.0 years [all results are indicated as mean ± standard error of the mean; *SEM*]; age range = 20 to 76 years; 26 female, 39 male) with (corrected to) normal visual accuracy, no colour-blindness, no signs of cognitive impairment (Mini-Mental-State

Examination score higher than 26; MMSE; Folstein et al., 1975) and no history of psychiatric or neurological diseases.

The current study’s data has been used previously (Overhoff et al., 2021). The same exclusion criteria regarding the neuropsychological assessment and the task performance were applied, resulting in the same subsample included in the analyses. In the previous publication, we thoroughly examined the metacognitive performance and its relation to behavioural parameters (response accuracy, response time, behavioural adjustments) as well as two electrophysiological potentials (i.e., the error/correct negativity, N_e, and the error/correct positivity, P_e; for detailed results and discussion thereof, see Overhoff et al., 2021). We did not report or analyse any response force measures in the previous publication.

The experiment was approved by the ethics committee of the German Psychological Society (DGPs). All participants gave written informed consent, and the study followed the Declaration of Helsinki.

2.2. Stimuli

The experiment consisted of a colour version of the Flanker task (Eriksen & Eriksen, 1974) with four response options intended to increase conflict and thereby the number of errors while ensuring feasibility for participants of all ages. Four target colours were mapped onto both hands’ index and middle fingers. In each trial, we presented one central, coloured target square flanked by two squares on the left and right side, respectively. Participants had to respond to the central target by pressing the corresponding finger. The flankers were presented slightly before the target appeared to increase their distracting effect. Flankers could be of the same colour as the target (congruent condition), of one of three additional neutral colours that were not mapped to any response (neutral condition), or of another target colour (incongruent condition).

2.3. Experimental Paradigm

Each trial started with the presentation of a white fixation cross on black background for 500 ms. The fixation cross was replaced by the two flankers, followed by the target after 50 ms and the two flankes and the target remained on screen for another 100 ms. Participants pressed their left or right index or middle finger to indicate their decision (see Figure 1B). Participants were instructed to respond as fast and accurately as possible. A black screen was presented until a response was registered (max. 1200 ms) and an additional 800 ms before presenting the confidence rating. For this rating, participants indicated their confidence in the decision on a four-point scale comprising the options 'surely wrong', 'maybe wrong', 'maybe correct', and 'surely correct' (max. 2000 ms). A jittered intertrial interval of 400 to 600 ms preceded the subsequent trial. If no response was registered in the decision task, the participants received feedback about being too slow, and the trial was terminated. The sequence of an experimental trial is depicted in Figure 1A.

2.4. Procedures

Prior to testing, we collected demographic details, and the participants conducted a brief colour discrimination test without any time pressure or cognitive load to ensure that they were capable of correctly discriminating the stimulus colours used in the experiment. The neuropsychological tests for assessing the exclusion criteria (BDI; MMSE; Edinburgh Handedness Inventory, EHI; Oldfield, 1971) were administered after the main experiment.

Participants first performed 18 practice trials without confidence rating, receiving feedback about their accuracy, which could be repeated if necessary. Two practice blocks of 72 trials without feedback followed. Another practice block then introduced the confidence rating. The actual experiment consisted of five blocks with 72 trials each, with optional breaks after each block. The electroencephalogram (EEG) was recorded throughout the testing session. Note that the EEG results have been reported in our previous publication (Overhoff et al., 2021).

2.5. Apparatus

The participants were seated in a noise-insulated and dimly lit testing booth at a viewing distance of 70 cm to the screen (LCD monitor, 60 Hz). A chin rest minimised non-task related movements.

For response recording, we used force sensitive keys with a sampling rate of 1024 Hz and a high temporal resolution that is superior to standard keyboards (Figure 1B; Stahl et al., 2020). The keys were calibrated to the fingers' weight before and during the experiment. The keys could be adjusted to the hand size, and a comfortable hand position was ensured by a wrist rest. An applied force was registered as a response when it exceeded a threshold of 40 cN.

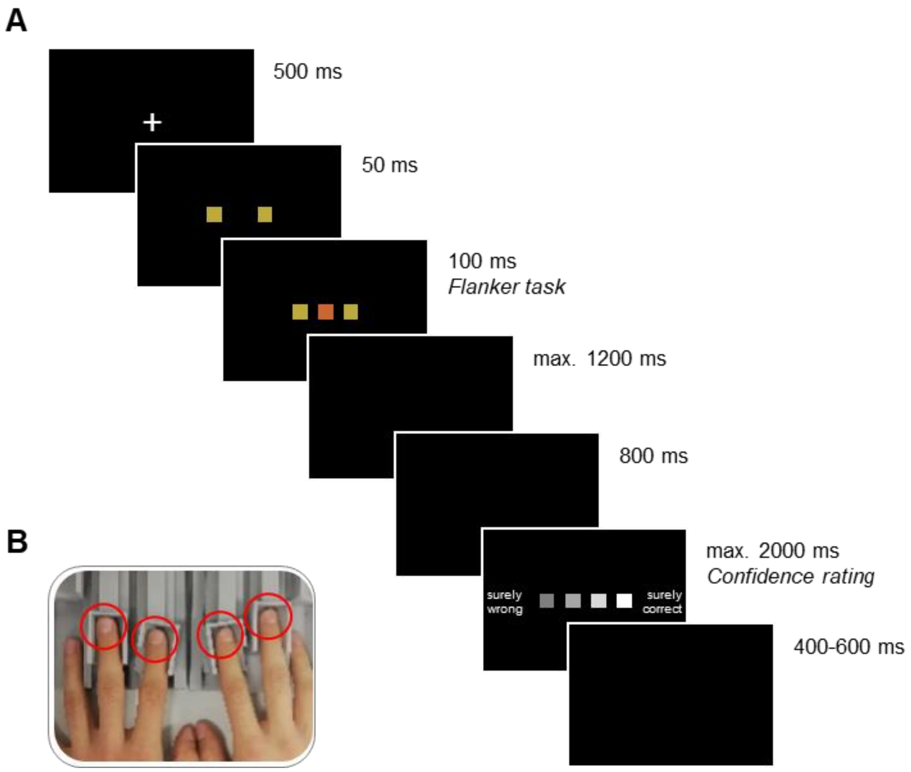


Figure 1. (A) Trial structure (here, incongruent condition illustrated). Each trial commenced with the presentation of a fixation cross. Subsequently, two coloured squares (flankers) were presented, and a third square (of the same or a different colour; target) was added shortly after. The stimuli disappeared after 100 ms and the ensuing black screen, where participants were instructed to make a response by pressing one of four response keys mapped onto one colour each, remained until a response was registered (maximum 1,200 ms). If no response was given, the German words for ‘too slow’ were shown, and the trial was terminated. Otherwise, after another black screen, the confidence rating scale was presented, which remained on the screen until a judgement (the four fingers were mapped onto the squares according to their spatial location) was made (maximum 2,000 ms). The next trial started after another black screen of random duration between 400 and 600 ms. (B) Force-sensitive response keys. Left and right index and middle fingers (red circles) were placed on adjustable finger rests.

The colour discrimination test was programmed using Presentation software (Neurobehavioural Systems, version 14.5) and the main task using uVariotest software (version 1.978).

2.6. Analysis

Response time (RT) was defined as the time from target stimulus presentation to the initial crossing of the response force threshold of 40 cN by any response key. Peak force (PF) was defined as the maximum of a force pulse following the crossing of the threshold. Additionally, we measured the time from response onset to the time of the PF (only used for the exclusion of trials).

We excluded from the analysis: invalid trials, which were too slow, responses without confidence rating, responses with an RT below 200 ms (indicating premature responding), a

PF below 40 cN (indicating incomplete or aborted responding), a time to PF of more than three standard deviations above the mean (indicating that the response was not of the expected ballistic nature), and recording artefacts (implausible time between response onset and time of the PF, incorrect identification of response key in case of multiple responses).

As a first step, to characterise the distribution of the behavioural parameters of interest independent of confidence, we computed paired samples *t*-tests at the group level to compare RT, PF, and their dispersion between correct and incorrect trials. For the investigation of age-related effects, we used a series of linear regressions with the predictor age for each of the following variables: error rates (ER; the proportion of valid responses that were incorrect), mean confidence ratings, mean RT and PF, and standard deviation of RT and PF. The latter analyses were performed separately for errors and correct responses.

Next, data were analysed using generalised linear mixed-effects models (GLMMs) with a beta distribution using the glmmTMB package (version 1.0.2.1; Brooks et al., 2017) in R (version 4.0.5; R Core Team, 2021). We chose this modelling approach because the beta distribution is assumed to better account for data that are not normally distributed and doubly bounded (i.e., having an upper and a lower bound; here: 1, “surely wrong”, and 4, “surely correct”), which applies to our confidence data (Verkuilen & Smithson, 2012). All continuous predictor variables were mean centred and scaled for model fitting, and confidence was scaled to the open interval (0,1; i.e., the range is slightly compressed to avoid boundary observations; Verkuilen and Smithson, 2012). Analyses were again conducted separately for correct responses and errors.

We examined the effects of age and the two parameters (RT, PF) of the response on confidence ratings using the following regression model structures (separately for the subsets of errors and correct responses):

(1) Confidence ~ Age + (RT | Participant)

(2) Confidence ~ RT*Age + (RT | Participant)

(3) Confidence ~ PF*Age + (RT | Participant)

(4) Confidence ~ RT*PF*Age + (RT | Participant)

RT and PF were used as fixed effects, and age was included as a covariate due to its documented negative effect on metacognitive accuracy (i.e., a negative effect on confidence for correct responses and a positive effect on confidence for errors; Overhoff et al., 2021; Palmer et al., 2014). For the most complex model, we considered an interaction term between all three factors, as RT and PF are known to vary across age (Dully et al., 2018), and previous work suggests potential interactions between RT and PF (Bode & Stahl, 2014; Gajdos et al., 2019). We fitted random intercepts for participants, allowing their mean confidence ratings to differ. If possible and the models converged, random slopes by participant were added for the predictors of interest to account for individual differences in the degree to which these were related to the confidence ratings (Barr et al., 2013). Models were checked for singularity and multicollinearity by calculating the variance inflation factor (VIF) using the performance package (version 0.7.2; Lüdecke et al., 2021).

We compared model fits including all effects of interest (model 4) to models including only one (models 2, 3) or no effect of interest (model 1) using likelihood ratio tests, and computed Wald *z*-tests to determine the significance of each coefficient. This means that, if a model including one predictor of interest (e.g., RT) fits the data better than a model including no effect of interest, this predictor has a relevant effect on confidence, and its inclusion in the model allows for a better prediction of participants' ratings.

To follow up on significant interaction effects between age and the predictors of interest, we calculated slopes for three values of age (the mean and one standard deviation above and below the mean). Additionally, for statistical analysis of the transitions between these values, we computed Johnson-Neyman intervals using an adapted version of the *johnson_neyman* function of the interactions package (version 1.1.0; Long, 2019). This analysis reveals whether the statistical effect of the response parameters on confidence is conditional on the entire range of the moderator age, or just a sub-range, thus providing bounds for where the observed interaction effect is significant.

3. Results

3.1. Overview of response parameters

On average, participants had an error rate of 15.4 ± 1.6 %. Correct trials had a mean RT of 709.2 ± 11.5 ms and were faster [$t(64) = -3.01$, $p = .004$] and had a smaller standard deviation [$t(64) = -5.53$, $p < .001$] than error trials with an RT of 734.3 ± 13.9 ms. The mean peak force (PF) was higher for correct trials (236.2 ± 13.1 cN) compared to errors [191.7 ± 10.3 cN; $t(64) = 5.51$, $p < .001$] but did not differ in its standard deviation [$t(64) = -0.21$, $p = .836$; see supplementary Figure S1].

3.2. Effect of age on response parameters

We have already reported the relationship between age and error rate, RT, and confidence in our previous publication (Overhoff et al., 2021) based on a slightly different subset of trials to the one used here (due to additional force-related exclusions of trials in this study). Our initial results were confirmed using a series of linear regression analyses, each using age as the predictor for one of the following variables: We found that, at group level, the error rate increased with age [$F(1,63) = 34.12$, $p < .001$, $\beta = 0.005$, $SE = 0.001$, $t = 5.84$]. RT increased with age for correct [$F(1,63) = 27.07$, $p < .001$, $\beta = 3.115$, $SE = 0.599$, $t = 5.20$]

and incorrect responses [$F(1,63) = 10.16$, $p = .002$, $\beta = 2.568$, $SE = 0.806$, $t = 3.19$], while age did not significantly predict PF for either type of response [correct: $F(1,63) = 0.02$, $p = .884$, $\beta = 0.120$, $SE = 0.816$, $t = 0.15$; error: $F(1,63) = 1.33$, $p = .254$, $\beta = 0.734$, $SE = 0.637$, $t = 1.15$]. RTs were more variable with higher age for correct responses [$F(1,63) = 7.43$, $p = .008$, $\beta = 0.477$, $SE = 0.175$, $t = 2.73$], but not errors [$F(1,63) = 0.04$, $p = .837$, $\beta = 0.053$, $SE = 0.255$, $t = 0.21$]. Similar to the mean PF, the standard deviation of PF did not change with age [correct: $F(1,63) = 0.15$, $p = .704$, $\beta = 0.164$, $SE = 0.429$, $t = 0.38$; error: $F(1,63) = 0.08$, $p = .774$, $\beta = 0.137$, $SE = 0.475$, $t = 0.30$]. These results are illustrated in the supplementary Figure S1.

The mean confidence (in the decision being correct, on a scale from 1 to 4) for correct responses (3.82 ± 0.02 for the entire sample) decreased with age [$F(1,63) = 22.42$, $p < .001$, $\beta = -0.007$, $SE = 0.002$, $t = -4.74$]. This finding indicates that the older participants were, the less confident they were in being correct when responding correctly. Contrarily, the mean confidence for errors (2.35 ± 0.08 for the entire sample) increased with age [$F(1,63) = 21.96$, $p < .001$, $\beta = 0.019$, $SE = 0.004$, $t = 4.69$; see supplementary Figure S2]. Hence, the older the participants were, the less sure they were that the decision was wrong when making an error. We have recently described this phenomenon as an age-related tendency to use the middle of the confidence scale, pointing towards increased uncertainty in older adults (Overhoff et al., 2021).

3.3. Modelling of confidence

Variance inflation factors across all models with interactions were < 2.03, indicating low collinearity (< 5; James et al., 2013) between the predictors, and the models were not overfitted, as the fits proved not to be singular.

3.3.1. Confidence in correct decisions

We computed likelihood ratio tests to compare the model fit of the winning model to the three other models. These tests revealed that for correct decisions, model 2 (i.e., Confidence ~ RT*Age + (RT | Participant)), which included the interaction between RT and age, fitted the data best. It was superior to model 1 (the null model), which included only the fixed effect of age [$\chi^2(2) = 38.15$, $p < .001$], and model 3, which included only the interaction between PF and age [$\chi^2(2) = 33.71$, $p < .001$]. Moreover, model 4, which included the full interaction between PF, RT and age, did not improve the fit further [$\chi^2(4) = 8.64$, $p = .071$].

The best fitting model showed significant negative effects of age [$\beta = -0.147$, $SE = 0.039$, $z = -3.75$, $p < .001$] and RT [$\beta = -0.109$, $SE = 0.018$, $z = -6.13$, $p < .001$] on confidence and a significant interaction between the two factors [$\beta = -0.050$, $SE = 0.017$, $z = -2.96$, $p = .003$; Table 1]. Given that we found a significant interaction between RT and age, we computed simple slopes for three values of age (the mean and one SD above and below the mean). The analysis revealed that the negative effect of RT on confidence (i.e., higher confidence for faster responses) increased with older age [Figure 2A; -1SD (i.e., younger adults): $\beta = -0.097$, $SE = 0.026$, $z = -3.75$, $p = .001$; mean (middle-aged adults): $\beta = -0.147$, $SE = 0.018$, $z = -8.32$, $p < .001$; +1SD (i.e., older adults): $\beta = -0.197$, $SE = 0.023$, $z = -8.53$, $p < .001$]. The Johnson-Neyman technique revealed that the effect of RT on confidence became significant from around 24 years of age onwards (higher bound of insignificant interaction effect: 24.41; Figure 2A).

3.3.2. Confidence in erroneous decisions

For errors, the best fitting model was model 3 (i.e., Confidence ~ PF*Age + (RT | Participant)), which included the interaction between PF and age. The likelihood ratio tests revealed that this model fitted the data better than model 1 (the null model) [$\chi^2(2) = 8.38$, $p = .015$] and model 2, which included the interaction between RT and age [$\chi^2(2) = 5.36$, $p <$

.001], and model 4, which included the interaction between all three factors, did not show an improved model fit, either [$\chi^2(4) = 4.47$, $p = .347$].

The winning model showed a significant effect of age [$\beta = 0.288$, $SE = 0.066$, $z = 4.40$, $p < .001$] and a significant interaction between PF and age [$\beta = 0.080$, $SE = 0.030$, $z = 2.67$, $p = .008$], but no main effect of PF [$\beta = -0.018$, $SE = 0.029$, $z = -0.65$, $p = .519$; Table 2]. To further unpack the interaction effect, we ran a simple slope analysis. This analysis showed a negative relationship between confidence and PF only for younger adults, while with increasing age, the slope was not significantly different from zero [Figure 2B: -1SD (i.e., younger adults): $\beta = -0.098$, $SE = 0.038$, $z = -2.60$, $p = .009$; mean (middle-aged adults): $\beta = -0.018$, $SE = 0.029$, $z = -0.645$, $p = .519$; +1SD (older adults): $\beta = 0.061$, $SE = 0.045$, $z = 1.37$, $p = .170$]. Computation of the Johnson-Neyman interval showed that above an age of about 44 years (lower bound of significant interaction effect: 43.501), PF was no longer significantly associated with confidence (Figure 2B).

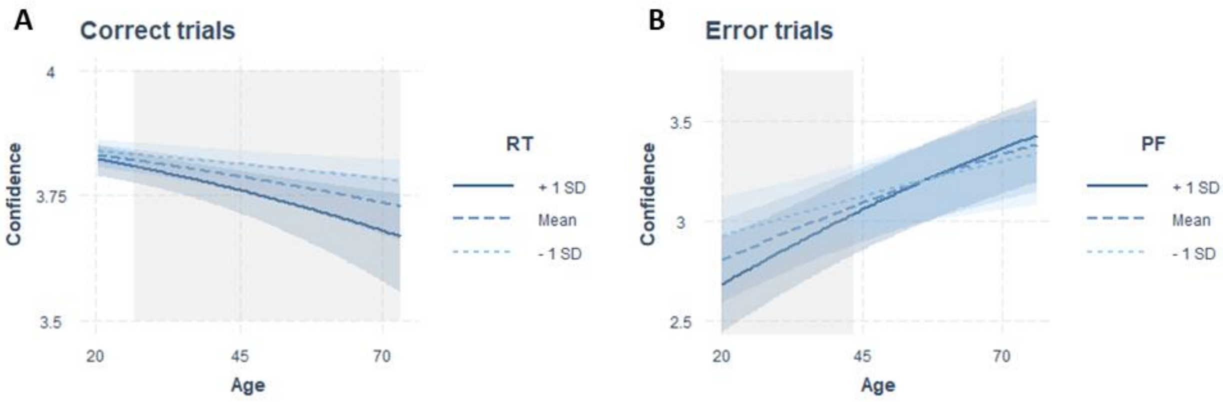


Figure 2. Interaction plots including the predictors of the models predicting confidence best for errors and correct responses. (A) Regression of age on confidence in correct trials with the moderator RT. (B) Regression of age on confidence in error trials with the moderator PF. Regressions are shown for the moderator fixed on the mean (dashed line) and one standard deviation above (solid line) and below (dotted line) the mean. Blue shaded areas indicate

confidence intervals, and grey shaded areas indicate the age range in which a significant effect of RT (in correct trials) or PF (in error trials) on confidence is observed, resulting in the significant interaction effect.

Table 1.
Regression coefficients (Estimate), standard errors (SE), and associated z- and p-values from the winning generalised linear (beta distribution) mixed-effects model for predicting confidence in correct responses.

Fixed effects	Estimate	SE	*z*	*p*
(Intercept)	2.630	0.044	59.75	**<.001**
RT	-0.109	0.018	-6.13	**<.001**
Age	-0.147	0.039	-3.75	**<.001**
RT*Age	-0.050	0.017	-2.96	**.003**

Table 2.
Regression coefficients (Estimate), standard errors (SE), and associated z- and p-values from the winning generalised linear (beta distribution) mixed-effects model for predicting confidence in incorrect responses.

Fixed effects	Estimate	SE	*z*	*p*
(Intercept)	-0.085	0.065	-1.31	.192
PF	-0.018	0.029	-0.65	.519
Age	0.288	0.066	4.40	**<.001**
PF*Age	0.080	0.030	2.67	**.008**

4. Discussion

This study investigated age-related changes in the relationship between the temporal and motor response parameters RT and PF with decision confidence. Overall, higher confidence was related to faster and less forceful responses. We could further show that, across the entire sample, confidence was associated with both parameters, and these effects were moderated by performance accuracy: While RT was related to confidence in correct responses, peak force was related to confidence in error trials. Finally, age interacted with the response parameters so that with higher age, the effect of RT on confidence was more pronounced, while the effect of PF on confidence was diminished.

We will first focus on the observed age-related associations between confidence and response parameters and subsequently discuss possible interpretations within two different theoretical frameworks.

4.1. Behavioural correlates of confidence

In a complex conflict task, we replicated one of the most robust findings on decision confidence, namely a negative relationship between RT and confidence (Fleming et al., 2010; Rahnev et al., 2020). In correct trials, the higher participants rated their confidence in a decision, the faster they had made the decision. In addition, we found a negative relationship between PF and confidence for errors (higher PF was related to lower confidence for the younger participants), which has not been reported before. Observing the latter association is interesting *per se* because participants' attention was not directed to the applied force in any way (i.e., participants were not aware of the PF assessment), whilst the relevance of speed had been stressed in the instructions. A recent study (Turner, Angdias, et al., 2021) showed, in a sample of young participants, that when higher levels of force had to be produced to report the (correct) decision, participants' confidence ratings were higher. While these results do not mirror ours, it should be noted that these findings also cannot be directly compared as their study was conceptually different to our study design and explicitly required participants to produce different force ranges. However, these studies together highlight the added value of assessing response force. In our study, the differential effect of RT and PF for correct and error trials, respectively, further highlights that these are dissociable parameters of a response, supporting a model of Ulrich and Wing (1991; see Armbrecht et al., 2013; Jaśkowski et al., 2000) that RT and RF do not reflect just two sides of the same coin. Further, our findings stress the significance of including incorrect responses as a distinct response type in the corresponding analyses.

Interestingly, the observed associations between the response parameters and confidence differed across the studied age range. While age-related changes in metacognitive performance, and error detection in particular, have been shown across tasks and domains (Harty et al., 2013; Niessen et al., 2017; Palmer et al., 2014), specific characteristics of confidence judgements have rarely been investigated in the context of healthy ageing. The current study revealed a stronger association with increasing age between confidence and RT in correct trials and a weaker association between confidence and PF in errors. Since the current sample of participants covered a broad age range, this constitutes a further step in identifying and understanding age-related changes in metacognitive performance.

We will present two complementary but not exclusive interpretations of the observed age-related variations in the following. In the first part, we attempt to explain our findings under the assumption that response characteristics simply co-occur with the build-up of confidence. In contrast, in the second part, we assume that response parameters comprise additional information about the decision accuracy that is integrated into confidence during its formation process.

4.2. Response parameters as the expression of confidence

One possible framework for explaining the experimental findings is to assume that the level of decision confidence is expressed in the RT or PF of the response indicating this decision, either because confidence defines the response parameters or because a common process drives both confidence and the two parameters. In other words, if a participant is highly confident in a decision, this will affect the speed and the force with which they report this decision. Research has identified multiple stimulus-related characteristics that alter the accuracy of confidence judgements, like relative and absolute evidence strength (Y. H. Ko et al., 2022; Peters et al., 2017) or evidence reliability (Boldt et al., 2017). If sensory evidence is unambiguous, an easy decision will accordingly lead to high certainty of having made a correct response. In turn, if the participant nevertheless responds incorrectly but changes their mind and detects this error, the certainty of having made an error will be high (i.e., resulting in a low confidence rating). It is intuitive to imagine that high *certainty* of having made a correct or incorrect response (which is identical to very high or very low confidence, respectively) will lead to fast and more forceful responses.

Notably, neither RT nor PF showed the expected pattern of change as would be expected if one or both parameters simply mirrored a decline in confidence with age. Arguably, it might still be possible to explain the differential interactions with age by assuming that multiple other sources (e.g., perception, attention, response selection, motor processes) cause the observed relationships between confidence and the two response parameters. If ageing impacts (some of) these sources differentially, this might result in altered associations between confidence, RT and PF, as observed here. For instance, a cognitive process that is differentially susceptible in older compared to younger adults might affect the RT-confidence relationship but spare the relationship between PF and confidence. However, as we did not systematically investigate these other processes in the present study, we can neither support nor rule out these assumptions.

4.3. Modulation of confidence by response parameters

Alternatively, our findings could also be interpreted in line with recent studies postulating that parameters of a response indicating a decision may serve as an additional source of evidence that is integrated into confidence judgements about this decision – especially in ambiguous situations (Filevich et al., 2020; Gajdos et al., 2019; Pereira et al., 2020; Turner, Angdias, et al., 2021; Wokke et al., 2020). These studies showed that confidence could be altered, for instance, by applying TMS to the dorsal premotor cortex or by instructing participants to move faster (Fleming et al., 2015; Palser et al., 2018). Using very different methodological approaches, these studies mutually indicate that the post-decisional evidence accumulation might incorporate response characteristics of the initial decision into the subsequent confidence rating. Although our study design assessing the relationship of confidence with RT and naturally occurring PF precludes any conclusions regarding the causal direction of effects, it is nevertheless interesting to reflect on our results within this framework.

Looking at the overall relationship between confidence and the two response parameters, our differential findings for errors and correct responses suggest serial processing. First, the RT-related information might be ‘read out’ by the monitoring system and serve as an interoceptive cue about the difficulty of a decision. This assumption is in line with previous work (Dotan et al., 2018; Fleming et al., 2010; Gajdos et al., 2019; Kiani et al., 2014; Rahnev et al., 2020). This interpretation would suggest that the decision-makers arrive at a higher confidence judgement because they also register having responded faster (e.g., via the efference copy (Latash, 2021) or the later representation of their action).

However, this proposed mechanism might exclusively operate in correct trials to refine confidence judgements. For the relationship between confidence and RT in error trials, which were on average slower than correct trials, it must be considered that a variety of aspects can cause errors (e.g., lack of attention, perceptual lapse), and the response profiles of errors are similarly heterogeneous. Therefore, in case of conflict (which is present in error trials), RT might no longer yield reliable information about the task requirements, and the monitoring system might probe PF instead as an alternative response parameter to compensate for the lack of reliable RT when computing confidence. In support, recent work indicated that within a similar speed range for responses, the PF in error trials was related to decision confidence (Stahl et al., 2020). Hence, while RT might not differentiate confidence levels in errors, variations in PF may well capture this information and could therefore be integrated into the final confidence judgement.

Given the frequently described decline of metacognitive abilities with older age, which was also shown in our previous analysis of the current data set (by using the *Phi* correlation coefficient (Nelson, 1984) for the analysis of metacognitive accuracy (Overhoff et al., 2021)), it seems likely that the older adults were lacking relevant input for the computation of confidence, making it harder for them to accurately rate their decisions. Consequently, one possibility is that the stronger association between confidence and RT in older adults might reflect a compensation mechanism. To explain, while our study does not allow for firm conclusions as to why metacognitive accuracy declined in older adults, it appears that the input to the performance monitoring system was diminished (or not adequate anymore) and did not allow for computing confidence with the same level of accuracy as in younger adults. Therefore, it is possible that the stronger reliance on RT (in correct trials) might reflect the attempt to compensate for this by relying more on other sources of input, like the interoceptive feedback about the response speed (Fleming et al., 2010; Palser et al., 2018). However, it remains unclear whether this compensation fails, as, despite more substantial reliance on RT information, confidence judgements were still poorer compared to younger adults. This could be plausible, for example, if the monitoring of response parameters itself might also become poorer with increasing age. Alternatively, it is also possible that this compensation was indeed (somewhat) successful, and without incorporating RT information more strongly, confidence judgements would be even worse. Ultimately, our study cannot resolve this question.

For error trials, we observed that the relationship between confidence and response force diminished with age. One explanation might be related to the finding that healthy ageing has been associated with diminished neural specificity for errors (Endrass, Schreiber, et al., 2012; Harty et al., 2017; Overhoff et al., 2021; Park et al., 2010), meaning that older adults might have generally been worse at detecting the errors in the first place. A recent fMRI study has extended these findings by showing that the activity related to error awareness was specifically reduced in older adults (Sim et al., 2020). Based on these findings, our results could be interpreted as another instance of an age-related error-specific processing deficit. This functional processing deficit might also extend to the sensorimotor feedback of the produced force. The read-out of the response force – which might be used to infer confidence in case of errors – might thus not be readily accessible by older adults and potentially contribute to the demonstrated deficits in metacognitive accuracy.

4.4. Limitations

While the simultaneous recording of two response parameters for each response constitutes a strength of the present study, treating RT and PF as equivalent may be problematic. We have discussed RT and PF as separate but comparable features of motor activity, even though their apparent relevance differed largely. Force was produced without constraints, while the time to report the decision was limited to 1,200 ms and exerted considerable time pressure on the participants. Therefore, it would be interesting to examine changes in the modulation of confidence by RT and PF without limiting the time to respond. Moreover, since the RT in a given trial represents the sum of the time for stimulus-related processes (between stimulus onset and the start of the response movement) and the time for motor-related processes (e.g., movement time – the time between starting and terminating a response movement), future studies should additionally assess movement time and its relation to confidence.

As mentioned above, this study cannot resolve the question of causality of the observed associations. Based on the described literature, it is reasonable to speculate that our findings can be explained within the framework of response dynamics informing confidence judgements. However, we have carefully outlined an alternative explanation and acknowledge that both lines of interpretation may be valid in part.

5. Conclusion

Corroborating recent evidence, we revealed significant associations between decision confidence and the parameters of the responses indicating this decision. Furthermore, we extended these findings by showing that confidence was associated with fine-grained changes in the time taken to report a decision and the force invested in this response. These relationships were moderated by the accuracy of the response, and, most importantly, changed markedly across the adult life span. This notion should encourage the recording of response force in behavioural experiments whenever possible, as it might uncover specific effects that cannot be revealed by measuring other response parameters, like response times. While a causal explanation of these findings was beyond the scope of this study, one possible interpretation is that the observed age-related changes in the pattern of associations reflect a mechanism in the computation of confidence and may even constitute one aspect of the frequently observed decline in metacognitive ability with older age.

Acknowledgements

We thank all colleagues from the Institute of Neuroscience and Medicine (INM-3), Cognitive Neuroscience, the Department of Individual Differences and Psychological Assessment, and the Decision Neuroscience Lab at the Melbourne School of Psychological Sciences for valuable discussions and their support.

Disclosure statement

The authors declare no conflict of interest.

References

Armbrecht, A.-S., Gibbons, H., Stahl, J., 2013. Effects of response force parameters on medial-frontal negativity. *PLoS One 8*, e54681. https://doi.org/10.1371/journal.pone.0054681

Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Learn. 68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bode, S., Stahl, J., 2014. Predicting errors from patterns of event-related potentials preceding an overt response. *Biol. Psychol. 103*, 357–369. https://doi.org/10.1016/j.biopsycho.2014.10.002

Boldt, A., de Gardelle, V., Yeung, N., 2017. The impact of evidence reliability on sensitivity and bias in decision confidence. *J. Exp. Psychol. Hum. Percept. Perform. 43*, 1520–1531. https://doi.org/10.1037/XHP0000404

Brooks, M.E., Kristensen, K., Van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Machler, M.;, Bolker, B.M., 2017. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *R J. 9*, 378–399. https://doi.org/10.3929/ethz-b-000240890

Bunce, D., Macdonald, S.W.S., Hultsch, D.F., 2004. Inconsistency in serial choice decision and motor reaction times dissociate in younger and older adults q. *Brain Cogn. 56*, 320–327. https://doi.org/10.1016/j.bandc.2004.08.006

Charles, L., Yeung, N., 2019. Dynamic sources of evidence supporting confidence judgments and error detection. *J. Exp. Psychol. Hum. Percept. Perform. 45*, 39–52. https://doi.org/10.1037/xhp0000583

Cohen, M.X., van Gaal, S., 2014. Subthreshold muscle twitches dissociate oscillatory neural signatures of conflicts from errors. *Neuroimage 86*, 503–513. https://doi.org/10.1016/j.neuroimage.2013.10.033

Dotan, D., Meyniel, F., Dehaene, S., 2018. On-line confidence monitoring during decision making. Cognition 171, 112–121. https://doi.org/10.1016/j.cognition.2017.11.001

Dully, J., McGovern, D.P., O'Connell, R.G., 2018. The impact of natural aging on computational and neural indices of perceptual decision making: A review. *Behav. Brain Res. 355*, 48–55. https://doi.org/10.1016/j.bbr.2018.02.001

Endrass, T., Schreiber, M., Kathmann, N., 2012. Speeding up older adults: Age-effects on error processing in speed and accuracy conditions. *Biol. Psychol. 89*, 426–432. https://doi.org/10.1016/j.biopsycho.2011.12.005

Eriksen, B.A., Eriksen, C.W., 1974. Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys. 16*, 143–149.

Falkenstein, M., Yordanova, J., Kolev, V., 2006. Effects of aging on slowing of motor-response generation. *Int. J. Psychophysiol. 59*, 22–29. https://doi.org/10.1016/J.IJPSYCHO.2005.08.004

Feuerriegel, D., Murphy, M., Konski, A., Mepani, V., Sun, J., Hester, R., Bode, S., 2021. Electrophysiological correlates of confidence differ across correct and erroneous perceptual decisions. *bioRxiv*. https://doi.org/https://doi.org/10.1101/2021.11.22.469610doi:

Ficarella, S.C., Rochet, N., Burle, B., 2019. Becoming aware of subliminal responses: An EEG/EMG study on partial error detection and correction in humans. *Cortex 120*, 443–456. https://doi.org/10.1016/j.cortex.2019.07.007

Filevich, E., Koß, C., Faivre, N., 2020. Response-related signals increase confidence but not metacognitive performance. *eNeuro 7*, 1–14. https://doi.org/10.1523/ENEURO.0326-19.2020

Fleming, S.M., Daw, N.D., 2017. Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychol. Rev. 124*, 91–114. https://doi.org/10.1037/REV0000045

Fleming, S.M., Dolan, R.J., 2012. The neural basis of metacognitive ability. *Philos. Trans. R. Soc. B Biol. Sci. 367*, 1338–1349. https://doi.org/10.1098/rstb.2011.0417

Fleming, S.M., Maniscalco, B., Ko, Y., Amendi, N., Ro, T., Lau, H., 2015. Action-Specific Disruption of Perceptual Confidence. *Psychol. Sci. 26*, 89–98. https://doi.org/10.1177/0956797614557697

Fleming, S.M., Weil, R.S., Nagy, Z., Dolan, R.J., Rees, G., 2010. Relating introspective accuracy to individual differences in brain structure. *Science 329*, 1541–1543. https://doi.org/10.1126/SCIENCE.1191883

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res. 12*, 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Franz, E.A., Miller, J., 2002. Effects of response readiness on reaction time and force output in people with Parkinson's disease. *Brain 125*, 1733–1750. https://doi.org/10.1093/brain/awf192

Gajdos, T., Fleming, S.M., Saez Garcia, M., Weindel, G., Davranche, K., 2019. Revealing subthreshold motor contributions to perceptual confidence. *Neurosci. Conscious. 2019*, 1–8. https://doi.org/10.1093/nc/niz001

Harty, S., Murphy, P.R., Robertson, I.H., O’Connell, R.G., 2017. Parsing the neural signatures of reduced error detection in older age. *Neuroimage 161*, 43–55. https://doi.org/10.1016/j.neuroimage.2017.08.032

Harty, S., O’Connell, R.G., Hester, R., Robertson, I.H., 2013. Older adults have diminished awareness of errors in the laboratory and daily life. *Psychol. Aging 28*, 1032–1041. https://doi.org/10.1037/a0033567

Hautzinger, M., 1991. The Beck Depression Inventory in clinical practice. *Nervenarzt 62*, 689–696.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An introduction to statistical learning: with applications in R. Springer, New York.

Jaśkowski, P., Van Der Lubbe, R.H.J., Wauschkuhn, B., Wascher, E., Verleger, R.R., 2000. The influence of time pressure and cue validity on response force in an S1-S2 paradigm. *Acta Psychol. (Amst). 105*, 89–105. https://doi.org/10.1016/S0001-6918(00)00046-9

Kiani, R., Corthell, L., Shadlen, M.N., 2014. Choice certainty is informed by both evidence and decision time. *Neuron 84*, 1329–1342. https://doi.org/10.1016/j.neuron.2014.12.015

Ko, Y.H., Feuerriegel, D., Turner, W., Overhoff, H., Niessen, E., Stahl, J., Hester, R., Fink, G.R., Weiss, P.H., Bode, S., 2022. Divergent effects of absolute evidence magnitude on decision accuracy and confidence in perceptual judgements. *Cognition 225*, 105125. https://doi.org/10.1016/j.cognition.2022.105125

Ko, Y.T., Alsford, T., Miller, J., 2012. Inhibitory effects on response force in the stop-signal paradigm. *J. Exp. Psychol. Hum. Percept. Perform. 38*, 465–477. https://doi.org/10.1037/a0027034

Latash, M., 2021. Efference copy in kinesthetic perception: a copy of what is it? *J. Neurophysiol. 125*, 1079–1094. https://doi.org/10.1152/jn.00545.2020

Long, J.A., 2019. interactions: Comprehensive, User-Friendly Toolkit for Probing Interactions.

Lüdecke, D., Ben-Shachar, M.S., Patil, I., Waggoner, P., Makowski, D., 2021. performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *J. Open Source Softw. 6*, 3139.

Nelson, T., 1984. A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychol. Bull. 95*, 109–133.

Niessen, E., Fink, G.R., Hoffmann, H.E.M.M., Weiss, P.H., Stahl, J., 2017. Error detection across the adult lifespan: Electrophysiological evidence for age-related deficits. *Neuroimage 152*, 517–529. https://doi.org/10.1016/j.neuroimage.2017.03.015

Oldfield, R.C., 1971. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia 9*, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Overhoff, H., Ko, Y.H., Feuerriegel, D., Fink, G.R., Stahl, J., Weiss, P.H., Bode, S., Niessen, E., 2021. Neural correlates of metacognition across the adult lifespan. *Neurobiol. Aging 108*, 34–46. https://doi.org/10.1016/j.neurobiolaging.2021.08.001

Palmer, E.C., David, A.S., Fleming, S.M., 2014. Effects of age on metacognitive efficiency. *Conscious. Cogn. 28*, 151–160. https://doi.org/10.1016/j.concog.2014.06.007

Palser, E.R., Fotopoulou, A., Kilner, J.M., 2018. Altering movement parameters disrupts metacognitive accuracy. *Conscious. Cogn. 57*, 33–40. https://doi.org/10.1016/j.concog.2017.11.005

Park, J., Carp, J., Hebrank, A., Park, D.C., Polk, T.A., 2010. Neural Specificity Predicts Fluid Processing Ability in Older Adults. *J. Neurosci. 30*, 9253–9259. https://doi.org/10.1523/JNEUROSCI.0853-10.2010

Pereira, M., Faivre, N., Iturrate, I., Wirthlin, M., Serafini, L., Martin, S., Desvachez, A., Blanke, O., de Ville, D. Van, del Millán, J.R., 2020. Disentangling the origins of confidence in speeded perceptual judgments through multimodal imaging. *Proc. Natl. Acad. Sci. U. S. A. 117*, 8382–8390. https://doi.org/10.1073/pnas.1918335117

Peters, M.A.K., Thesen, T., Ko, Y.D., Maniscalco, B., Carlson, C., Davidson, M., Doyle, W., Kuzniecky, R., Devinsky, O., Halgren, E., Lau, H., 2017. Perceptual confidence neglects decision-incongruent evidence in the brain. *Nat. Hum. Behav. 1*, 1–21. https://doi.org/10.1038/s41562-017-0139

Quandt, F., Bönstrup, M., Schulz, R., Timmermann, J.E., Zimerman, M., Nolte, G., Hummel, F.C., 2016. Spectral Variability in the Aged Brain during Fine Motor Control. *Front. Aging Neurosci. 8*, 305. https://doi.org/10.3389/FNAGI.2016.00305

R Core Team, 2021. R: A language and environment for statistical computing.

Rahnev, D., Desender, K., Lee, A.L.F., Adler, W.T., Aguilar-Lleyda, D., Akdoğan, B., Arbuzova, P., Atlas, L.Y., Balcı, F., Bang, J.W., Bègue, I., Birney, D.P., Brady, T.F., Calder-Travis, J., Chetverikov, A., Clark, T.K., Davranche, K., Denison, R.N., Dildine, T.C., Double, K.S., Duyan, Y.A., Faivre, N., Fallow, K., Filevich, E., Gajdos, T., Gallagher, R.M., de Gardelle, V., Gherman, S., Haddara, N., Hainguerlot, M., Hsu, T.Y., Hu, X., Iturrate, I., Jaquiery, M., Kantner, J., Koculak, M., Konishi, M., Koß, C., Kvam, P.D., Kwok, S.C., Lebreton, M., Lempert, K.M., Ming Lo, C., Luo, L., Maniscalco, B., Martin, A., Massoni, S., Matthews, J., Mazancieux, A., Merfeld, D.M., O'Hora, D., Palser, E.R., Paulewicz, B., Pereira, M., Peters, C., Philiastides, M.G., Pfuhl, G., Prieto, F., Rausch, M., Recht, S., Reyes, G., Rouault, M., Sackur, J., Sadeghi, S., Samaha, J., Seow, T.X.F., Shekhar, M., Sherman, M.T., Siedlecka, M., Skóra, Z., Song, C., Soto, D., Sun, S., van Boxtel, J.J.A., Wang, S., Weidemann, C.T., Weindel, G., Wierzchoń, M., Xu, X., Ye, Q., Yeon, J., Zou, F., Zylberberg, A., 2020. The Confidence Database. *Nat. Hum. Behav. 4*, 317–325. https://doi.org/10.1038/s41562-019-0813-1

Sailer, A., Dichgans, J., Gerloff, C., 2000. The influence of normal aging on the cortical processing of a simple motor task. *Neurology 55*, 979–985. https://doi.org/10.1212/WNL.55.7.979

Salthouse, T.A., 2000. Aging and measures of processing speed. *Biol. Psychol. 54*, 35–54. https://doi.org/10.1016/S0301-0511(00)00052-1

Siedlecka, M., Koculak, M., Paulewicz, B., 2021. Confidence in action: Differences between perceived accuracy of decision and motor response. *Psychon. Bull. Rev. 28*, 1698–1706. https://doi.org/10.3758/s13423-021-01913-0

Sim, J., Brown, F., O'Connell, R., Hester, R., 2020. Impaired error awareness in healthy older adults: an age group comparison study. *Neurobiol. Aging 96*, 58–67. https://doi.org/10.1016/j.neurobiolaging.2020.08.001

Stahl, J., Mattes, A., Hundrieser, M., Kummer, K., Mück, M., Niessen, E., Porth, E., Siswandari, Y., Wolters, P., Dummel, S., 2020. Neural correlates of error detection during complex response selection: Introduction of a novel eight-alternative response task. *Biol. Psychol. 156*, 107969. https://doi.org/10.1016/j.biopsycho.2020.107969

Stahl, J., Rammsayer, T.H., 2005. Accessory stimulation in the time course of visuomotor information processing: Stimulus intensity effects on reaction time and response force. *Acta Psychol. (Amst). 120*, 1–18. https://doi.org/10.1016/j.actpsy.2005.02.003

Turner, W., Angdias, R., Feuerriegel, D., Chong, T.T., Hester, R., Bode, S., 2021a. Perceptual decision confidence is sensitive to forgone physical effort expenditure. *Cognition 207*, 104525.

Turner, W., Feuerriegel, D., Andrejević, M., Hester, R., Bode, S., 2021b. Perceptual change-of-mind decisions are sensitive to absolute evidence magnitude. *Cogn. Psychol. 124*, 101358. https://doi.org/10.1016/j.cogpsych.2020.101358

Ulrich, R., Wing, A.M., 1991. A Recruitment Theory of Force-Time Relations in thte Production of Brief Force Pulses: The Parallel Force Unit Model. *Psychol. Rev. 98*, 268–294.

Van Der Lubbe, R.H.J., Verleger, R., der Lubbe, R.H.J., Verleger, R., 2002. Aging and the Simon task. *Psychophysiology 39*, 100–110.

Verkuilen, J., Smithson, M., 2012. Mixed and mixture regression models for continuous bounded responses using the beta distribution. *J. Educ. Behav. Stat. 37*, 82–113. https://doi.org/10.3102/1076998610396895

Vickers, D., Packer, J., 1982. Effects of alternating set for speed or accuracy on response time, accuracy and confidence in a unidimensional discrimination task. *Acta Psychol. (Amst). 50*, 179–197. https://doi.org/10.1016/0001-6918(82)90006-3

Wessel, J.R., Dolan, K.A., Hollingworth, A., 2018. A blunted phasic autonomic response to errors indexes age-related deficits in error awareness. *Neurobiol. Aging 71*, 13–20. https://doi.org/10.1016/j.neurobiolaging.2018.06.019

Wokke, M.E., Achoui, D., Cleeremans, A., 2020. Action information contributes to metacognitive decision-making. *Sci. Rep. 10*, 1–15. https://doi.org/10.1038/s41598-020-60382-y

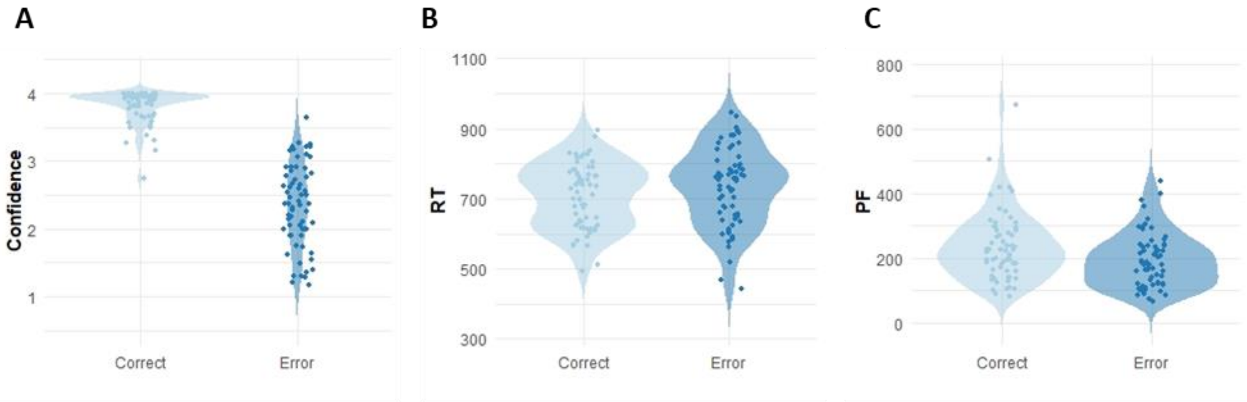
Yeung, N., Summerfield, C., 2012. Metacognition in human decision-making: Confidence and error monitoring. *Philos. Trans. R. Soc. B Biol. Sci. 367*, 1310–1321. https://doi.org/10.1098/rstb.2011.0416

Yordanova, J., Kolev, V., Hohnsbein, J., Falkenstein, M., 2004. Sensorimotor slowing with ageing is mediated by a functional dysregulation of motor-generation processes: evidence from high-resolution event-related potentials. *Brain 127,* 351–362. https://doi.org/10.1093/brain/awh042

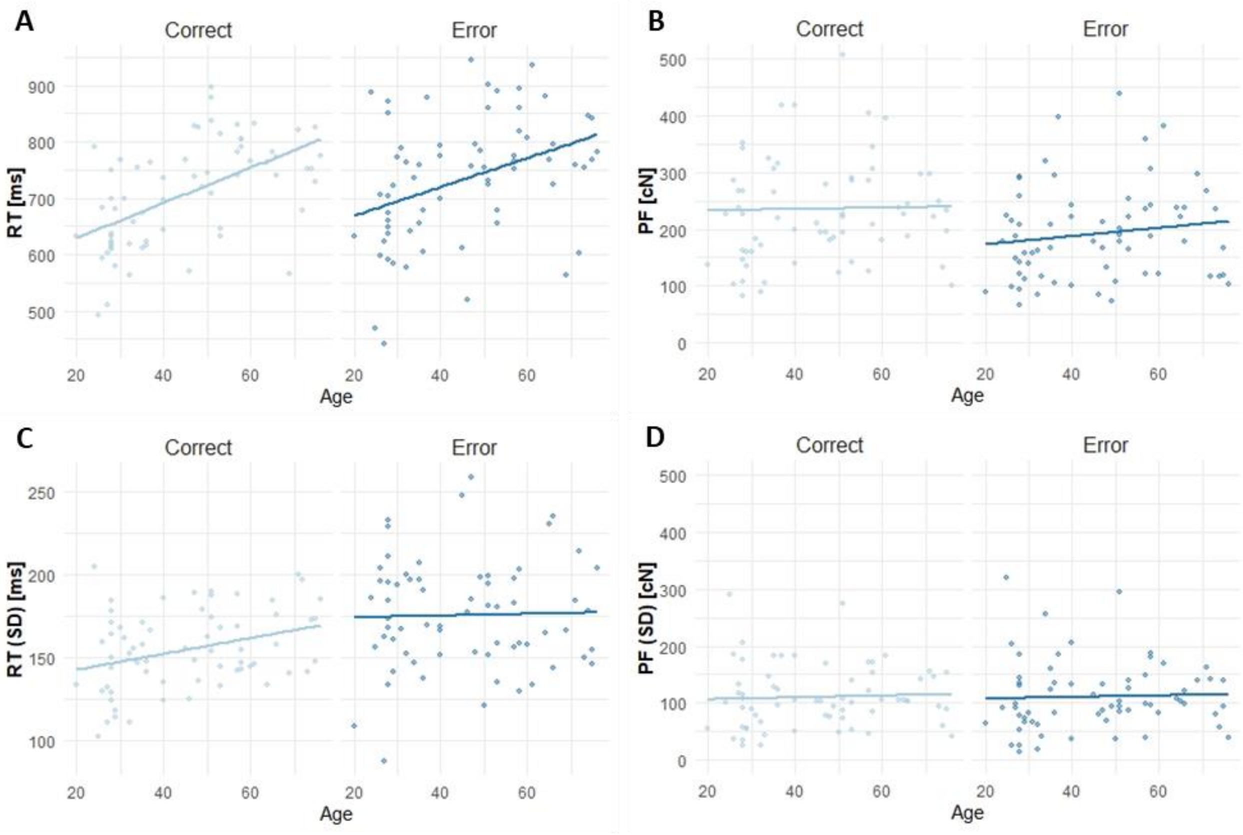
3.2.3. Supplement

S1. Figures control analyses

In order to get an overview of the distribution of the behavioural parameters of interest independent of confidence, we compared confidence ratings, RT, and PF between errors and correct responses. We assessed the effect of age on the movement parameters RT and PF. Results of *t*-tests and simple linear regressions are reported in the manuscript. Here, we are additionally providing the respective plots for illustration purposes.



Supplementary Figure S1. Distribution of (A) confidence ratings, (B) RT, and (C) PF for errors and correct responses. Dots indicate individual means (for confidence) or medians (for RT and PF), respectively.



Supplementary Figure S2. Regression of RT (left column) and PF (right column) on age for correct and error trials, respectively. Dots and fitted lines in the upper row indicate individual median RT/PF, and dots and fitted lines in the lower row indicate individual standard deviation (*SD*) for RT/PF. Median RT of correct and error trials and *SD* RT of correct trials increase significantly with age.

3.2.4. Summary

In this manuscript, we showed that the RT and PF of a first-order decision were related to the subsequent confidence judgement. Specifically, correct responses were given faster when they were rated with higher confidence and errors were reported with higher force when participants had low confidence that their decision was correct. Importantly, these patterns differed across the lifespan: The negative relationship between confidence and RT was even stronger in older adults, whereas confidence was only significantly related to PF in errors of younger adults. Older adults did not show any association between confidence ratings and the force of the prior response.

In the context of this thesis, we found an impact of age on behavioural correlates of decision confidence just as we did for the electrophysiological correlates in Study 1. Reported confidence was related to fine-grained variations in the response dynamics of the initial decision. As such, these findings are in line with studies suggesting that action-related information may contribute to the computation of confidence.

4. General Discussion

The central aim of this thesis was to better understand the development of metacognitive performance across the lifespan and how this is reflected in behavioural and neural correlates of decision confidence. To answer these questions, we developed a new paradigm assessing metacognitive accuracy that combines methods from the fields of decision confidence and error monitoring, and recorded response time, response force, and electrophysiological activity related to participants' decisions. We posed three specific research questions that were addressed across two studies. First, in Study 1a, we computed the accuracy of metacognitive judgements reported on a confidence scale in 65 adults (20 to 76 years) to investigate how metacognitive performance was affected by healthy ageing (Chapter 3, Section 3.1). Second, we explored two specific aspects that might be related to the observed age-related decline in metacognition performance. For this, we quantified how confidence was reflected in the $N_{e/c}$ and $P_{e/c}$ amplitudes of the EEG in Study 1b (Chapter 3, Section 3.1) and in the response time and force of decisions in Study 2 (Chapter 3, Section 3.2), and implications for underlying mechanisms of metacognition were discussed. Lastly, in order to define consequences of age-related declines in metacognition, we studied the effect of confidence judgements on adjustments of response caution in subsequent decisions in Study 1a (Chapter 3, Section 3.1).

In this final chapter, I will first review the key findings of the presented studies by focussing on relevant aspects in more detail. Then, I will discuss how our findings contribute to understanding the functional mechanisms of age-related metacognitive impairments. Lastly, I will take a broader perspective on short- and long-term implications of our findings and how they might translate to real life situations.

4.1. Review of key findings

4.1.1. Study 1a

In Study 1a, we investigated how metacognitive performance changed with age and how this affected the adaptation of decision policies. In a modified flanker task that was specifically designed to produce a high number of errors in order to assess signatures of confidence across correct and incorrect trials, participants made a decision and subsequently rated their confidence in this decision on a four-point scale comprising error detection (i.e., ‘surely wrong’) and full certainty in the accuracy of the decision (i.e., ‘surely correct’) at the extremes and uncertain options in the middle (i.e., ‘maybe wrong’, ‘maybe correct’). We tested a large sample of healthy adults throughout the lifespan. In contrast to the common use of two (or more) age groups, this provides the opportunity to investigate the trajectory of age-related changes of cognitive processes rather than only snapshots at two rather arbitrary cut-off values.

As expected, older adults responded slower than younger adults, but this slowing was not sufficient for the older adults to achieve comparable task performance as the error rate gradually increased with age (Salthouse et al., 1979). It should be noted that we intentionally refrained from implementing an adaptive algorithm that aligns task performance across participants (as e.g., Palmer et al., 2014; McWilliams et al., 2022) and expected to find systematic differences in task performance as a function of age. As our primary interest was in the second-order performance, we aimed to explore its correlates under natural conditions, that is, in a context where also first-order task performance changes across age.

4.1.1.1. Measuring metacognitive accuracy

In order to quantify metacognitive accuracy, we computed the Phi correlation coefficient (Kornell, Son, & Terrace, 2007; Nelson, 1984). This measure calculates for each trial the agreement between accuracy and the given confidence rating (i.e., a high confidence rating in a correct trial results in a high value of Phi and a low confidence rating in a correct trial leads to a low value of Phi) and averages these scores within each participant. Phi is affected by confidence bias, that is, when participants tend to give very high or very low confidence ratings in general (assessed on a half-scale of confidence and thus equal to very high or low certainty in being correct), Phi will underestimate these participants’ metacognitive accuracy (Shekhar & Rahnev, 2020). Since in our study younger adults tended towards high certainty ratings (i.e., very high or very low confidence) and older adults towards low certainty ratings (i.e., confidence ratings in the middle of the scale), the absolute values of Phi might be lower for both extremes, but the relative change in metacognitive accuracy across the studied age range would nevertheless be informative. In order to test whether age-related changes in metacognition were simply due to changes in task performance, we computed multiple linear regressions including the factor accuracy. This was chosen as a model-free alternative to the prevalent use of metacognitive efficiency, a measure based on signal detection theory for two-alternative forced-choice tasks that controls for task performance (Maniscalco & Lau, 2012).

4.1.1.2. Metacognitive performance across the adult lifespan

Phi decreased across the age range of our studied sample, indicating a decrease in metacognitive accuracy with older age. This pattern is consistent with the result of a previous study investigating changes in decision confidence across the lifespan (Palmer et al., 2014). In this study, participants made a two-alternative forced-choice perceptual judgement regarding which of two sets of stimuli contained a pop-out Gabor patch and rated their confidence in this decision. Notably, task performance was experimentally adjusted to around 70% accuracy across the entire sample and the confidence scale ranged from uncertainty to full certainty in being correct, thus, not allowing to indicate error detection. The authors used the metacognitive efficiency measure meta-d´/d´ that gives a relative measure of metacognitive performance independent of task performance (Maniscalco & Lau, 2012; Shekhar & Rahnev, 2020). While this measure was not applicable in our four-choice study design, we confirmed by using multiple linear regression that the age-related decline in metacognitive accuracy went beyond the observed age-related decline in task performance. Similar to Phi, metacognitive efficiency is affected by confidence bias, but in this case, it rather underestimates the performance of participants with a high confidence criterion (and thus a bias towards lower certainty; Shekhar and Rahnev, 2020). Thus, while Palmer and colleagues (2014) might have overestimated the age-related decline in metacognitive performance, the combination with our findings points towards a robust decline with increasing age. In sum, despite substantial differences in task design (primary task, confidence rating scale), sample (the mean age was about nine years higher in our study), and analysis approach (measure of metacognitive performance), we replicated the finding of a gradual decline in metacognitive performance across the adult lifespan by Palmer and colleagues (2014) and extended it from a classical perceptual discrimination task to a complex, four-choice conflict task typically used in studies of error monitoring.

4.1.1.3. Increased uncertainty in older adults

Upon closer inspection of the distribution of confidence ratings, we found that the age-related decline in metacognitive accuracy was driven by a differential use of the scale. Younger participants most often rated correct responses as ‘surely correct’ and errors as ‘surely wrong’, whereas older participants used the ratings indicating uncertainty (‘maybe correct’, ‘maybe wrong’) relatively more often. This finding contrasts with previous studies examining metacognitive evaluations across age. Paralleling findings from meta-memory research (Dodson et al., 2007; Hansson et al., 2008; Pansky et al., 2009) and experiments with patients with Alzheimer’s disease (Cosentino et al., 2007), Harty and colleagues (2013) assessed failures in daily life and showed that older adults largely overestimated the own abilities in real-world situations. In this study, metacognition was conceptualised as self-awareness and was assessed in several questionnaires of attentional control and memory functioning. Compared to statements of informants, younger participants tended to underestimate their abilities, while older participants tended to overestimate them, displaying decreased generalised self-awareness. Another study emphasised the danger of such overestimation by showing that older adults rated their own driving skills higher than objective measures indicated (Ross et al., 2012). Possible reasons for the increased uncertainty of older adults in this study are discussed in the following Section 4.2.1.

The age-related decline in metacognitive accuracy in our study was related to decreased confidence in correct responses and increased confidence in errors, which bridges the gap between studies of error monitoring and decision confidence. Mirroring our observation that older adults more often rated errors as ‘maybe correct’ compared to younger adults, studies on error monitoring frequently reported lower error detection rates in older adults (Harty et al., 2017; Niessen et al., 2017; Rabbitt, 1990; Sim et al., 2020). Notably, all these studies used an error awareness button for the metacognitive assessment, which is particularly prone to biases in confidence levels (similar to binary error detection ratings; Shekhar and Rahnev, 2020). To illustrate this, when a participant has a low confidence criterion, thus tending to report high confidence, they might frequently decide not to signal an error even though they might have some doubt in their decision. Accordingly, the use of a multiple-point confidence scale as in our experiment is more robust towards such biases as it likely covers a broader range of confidence criteria (Shekhar & Rahnev, 2020).

4.1.1.4. Adaptation of behaviour

Next, we aimed to understand whether interindividual differences in metacognitive accuracy affected future behaviour. To address this question, we examined how a given level of confidence was related to adjustments of subsequent behaviour and how this changed across the lifespan. Desender and colleagues (2019) argued that the internal sense of confidence might replace external feedback about the decision accuracy when this is absent as a signal for the regulation of decision policies (as it often is in real life). Supporting this view, we found that participants responded more cautiously after trials in which they indicated uncertainty or low confidence in their response.

Previous studies have provided ample evidence of behavioural adjustments after errors (for a review, see Danielmeier and Ullsperger, 2011). The most commonly investigated type of adjustment is post-error slowing (PES), a relative slowing of response times after errors compared to correct trials observed across various tasks (Notebaert et al., 2009; Rabbitt, 1966). The functional role of PES has been discussed to be either adaptive as it improves behaviour by slowing down parts of the decision-making process, or maladaptive, expressing momentarily impaired behaviour after errors due to limited resources that reflect a reorientation of attention (Wessel, 2018). Crucially, PES does not always improve the behavioural outcome, which might be related to the awareness of error commission (Danielmeier & Ullsperger, 2011). Although findings are again not unequivocal, growing evidence suggests that PES is larger after detected errors compared to undetected errors (Nieuwenhuis et al., 2001; Stahl et al., 2020; Wessel et al., 2018). The results of our study showed increased response caution after errors. However, response caution was also increased after responses where participants felt low confidence or uncertainty, independent of the objective accuracy. Since accuracy and confidence are closely related (e.g., Peters, 2022) and we cannot disentangle their separate influence with our study design, it is thus possible that confidence and not objective accuracy was the factor driving the adaptation of behaviour. In line with this, a recent study provided strong evidence for a causal effect of decision confidence on subsequent adjustments of behaviour (Desender et al., 2018). In a perceptual discrimination task, participants made an initial decision and had the option to choose to see the stimulus again before committing to a final decision and rating their confidence. The authors constructed two conditions that were matched for objective accuracy but differed in mean confidence ratings. It was shown that participants sought additional stimulus information more often in the condition with lower mean confidence, suggesting that this decision (i.e., the adjustment of behaviour towards seeking additional information) was based on the subjective sense of accuracy rather than the objectively observed accuracy. Accordingly, it could be argued that confidence (and not objective accuracy) might in fact predict subsequent response caution, for instance by slowing down in order to seek additional information after a low confidence response.

With regards to changes across the lifespan, we did not find evidence suggesting differences between younger and older adults in how their response caution was related to the accuracy or reported confidence of the previous response. A few studies have examined the effect of healthy ageing on post-error adjustments. Initially, Dutilh and colleagues (2013) modelled sources of increased PES in older adults and showed that the slowing was in index of a more cautious response strategy, but also less efficient information processing after an error. In contrast, Masina and colleagues (2018) investigated PES across different age groups and found that it was stable across the lifespan(Larson et al., 2016; Niessen et al., 2017)(Larson et al., 2016; Niessen et al., 2017)(Larson et al., 2016; Niessen et al., 2017). According to their interpretation, PES might constitute a mechanism compensating for general declines in performance monitoring that is not affected by ageing. Adjustments of behaviour in relation to the metacognitive evaluation of the preceding response have been investigated by Niessen and colleagues (2017) who found a slight decrease in PES after detected errors with older age. However, this effect failed to reach significance and could only be compared to correct responses but not undetected errors due to an insufficient number of trials of this response type. Again, the effect of metacognitive evaluations on behavioural adjustments in the context of ageing has not yet been investigated using confidence judgements. Since we showed a modulation of response caution by confidence across the lifespan, PES might also be related to variations in decision confidence rather than objectively observed performance or a binary classification of error detection, which could explain inconsistencies regarding the effect of age in previous studies. Taken together, we concluded that, in contrast to their strong decline in metacognitive accuracy, older adults were equally able as younger adults to implement adaptive trial-by-trial adjustments of response caution depending on their perceived accuracy of the prior decision.

4.1.1.5. Limitations

Overall, the implementation of our novel paradigm was successful as it was feasible for adults of all ages included in the final sample, it produced a reasonable number of errors, and participants used the entire range of the confidence scale. Despite these benefits, it has some shortcomings. First, although all participants made at least a few errors, a higher error rate would have increased the power for statistical analyses on errors. This could be achieved, for instance, by increasing the conflict inducing effect of the flankers (e.g., by presenting them even earlier or larger) or reducing the response deadline. A shorter presentation of the stimuli, in contrast, might affect additional cognitive processes (e.g., sensory encoding) that we intended to keep constant (Di Gregorio et al., 2018). At the same time, however, it should be considered that the task already posed higher demands on the older adults, as reflected in higher error rates and in a substantial number of datasets, primarily of older adults, which had to be excluded from analysis due to too many errors. Another possibility would be to adjust the task difficulty to the individual performance using an adaptive algorithm, which additionally allows to directly compare metacognitive accuracy between participants. Still, as explained above, we refrained from doing so in this first application of the paradigm.

Although participants indicated different degrees of certainty in being correct and in having made an error, the use of the confidence scale was skewed towards high confidence. This compelled us to collapse two confidence levels into one for analysis, thereby losing complexity of the data. When developing the paradigm further, it could be considered to use a continuous confidence scale and a slider with different starting positions (see e.g., Filevich et al., 2020).

With regards to the effect of confidence on behavioural adaptation, it should be noted that by modelling their data using the drift diffusion model (Ratcliff & McKoon, 2008), Desender and colleagues (2019) could show that increased response caution after low confidence decisions was related to increased boundary separation, that is, participants adopted higher internal decision thresholds for committing to a response. As we did not model our data, slowing after low confidence ratings could also be explained by a lower drift rate, for example, if a loss of attentional focus led to less efficient evidence accumulation in the following trial. However, increased cautiousness was also reflected in more accurate responses, which rather suggests a shift in decision threshold setting.

4.1.1.6. Conclusion

To reiterate our main conclusions from Study 1a, the combination of theoretical approaches from the fields of error monitoring and decision confidence allowed us to generalise the finding of a marked decline in metacognitive performance that accompanies ageing to another research field using a new paradigm. Moreover, the use of a confidence scale for the assessment of metacognitive judgements extended previous findings of reduced error awareness in older adults by relating them to increased uncertainty with older age (Harty et al., 2017; Niessen et al., 2017; Sim et al., 2020). Lastly, the adaptive adjustment of decision policies was found to be predicted by the subjectively perceived rather than the objective accuracy of prior decisions and was preserved across the lifespan.

4.1.2. Study 1b

In Study 1b, we explored how age-related changes in metacognition are reflected at the neural level. For this, we investigated the modulation of two established ERP correlates of error monitoring (i.e., $N_{e/c}$ and $P_{e/c}$) by confidence judgements. The study comprises the EEG element of Study 1. EEG recordings are particularly well suited for the study of decision-making and its metacognitive evaluation as they measure changes in neural activity with a high temporal resolution and thus, allow to track the time course of cognitive processes and isolate distinct functions that are closely aligned in time.

4.1.2.1. Modelling single-trial ERP activity

We focussed our analyses on the $N_{e/c}$ and the $P_{e/c}$ components. Typically, these components are averaged in time across the conditions of interest and across participants or, if the interest is in individual differences, within each participant. However, the variability in ERP latency and amplitude (and fluctuations across the experiment, e.g., in attention) gets lost in across-trial averaging. Thus, in order to model the relationship between confidence and EEG activity at a single-trial level, we measured ERP amplitudes in single-trial waveforms within participants. Due to the relatively low signal-to-noise ratio of the EEG signal it is not trivial to define the peak of a component at the level of single trials (Luck, 2014). As an approximation, we first computed the latency of the grand average peak of errors and correct responses, and then extracted adaptive mean amplitudes (100 ms interval surrounding the peak latencies) instead of peak ERP amplitudes from the EEG data, because they are more robust and less affected by noise (Clayson et al., 2013).

The analysis of ERP components at the single-trial level also encouraged us to apply a current source density (CSD) transformation to the EEG signal during preprocessing. This analysis leads to a clearer, reference-free separation of ERP components, because it serves as a spatial high-pass filter by removing contributions of temporally overlapping components of different neural generators (Kayser & Tenke, 2015; Luck, 2014). For the analysis of the ERP data, we used linear mixed-effects models, which account for the multi-level structure of our data (i.e., single-trial confidence ratings were nested within participants). The use of this method enhanced the sensitivity and reduced the noise in our estimates as it allows the inclusion of all participants and all trials in contrast to average waveforms that require a minimum number of trials per participant to be computed (Luck, 2014; Steele et al., 2016). Together, using linear mixed-effects models for single-trial ERP amplitudes allowed us to model subtle changes in neural activity with confidence across participants of different ages.

We found that both $N_{e/c}$ and $P_{e/c}$ showed the expected larger amplitudes for errors compared to correct responses. A novel finding of Study 1b was that the N_e and P_e amplitudes of errors were larger for lower confidence ratings and that this pattern was differentially affected by ageing. With older age, the N_e amplitude of errors decreased, and this was due to a particular decline in errors rated as ‘surely wrong’, while surprisingly, the P_e amplitude did not show the hypothesised decline with older age.

4.1.2.2. Effects of ageing on $N_{e/c}$ amplitudes

A decrease in the N_e amplitude with increasing age and the resulting smaller N_e - N_c difference have frequently been reported in studies of error monitoring (Endrass, Schreiber, et al., 2012; Falkenstein et al., 2001; Kolev et al., 2005; Pietschmann et al., 2011), while no effect of ageing was reported in very easy tasks (Larson et al., 2016; Niessen et al., 2017). Two studies that assessed effects of ageing on the N_e amplitude of detected and undetected errors revealed no interactions between age and error detection (Harty et al., 2017; Niessen et al., 2017). In contrast, Rausch and colleagues (2020) argued that the $N_{e/c}$ *does* reflect error detection after all. In young adults, they used a masking paradigm with varying stimulus-onset asynchronies and found a difference between N_e and N_c only in the easiest condition.

This supports the assumption that an $N_{e/c}$ can only emerge when the participant is aware of the actual correct response (Charles et al., 2013; Di Gregorio et al., 2018). While a previous study using response sets of different sizes showed reduced differences between N_e and N_c amplitudes for larger response sets (Maier et al., 2010), we found significantly larger N_e amplitudes, suggesting that our paradigm with more than the common two response options did not inhibit early error monitoring processes or the adjustment of future behaviour. However, it might have indeed been more difficult for older adults to build a representation of the correct response during the decision-making process. Alternatively, it is also possible that the analysis of the relationship between confidence judgements and ERP amplitudes at the single-trial level in our study was more informative regarding subtle variations in the $N_{e/c}$ amplitude with the metacognitive evaluation than averaged waveforms measured in previous studies.

4.1.2.3. Effects of ageing on $P_{e/c}$ amplitudes

The P_e, on the other hand, was not affected by ageing in our study, while Harty and colleagues (2017) and Niessen and colleagues (2017) found that the increased amplitude for detected compared to undetected errors was markedly reduced in older adults. In keeping with the argumentation of Rausch and colleagues (2020), the fact that we observed a similar strong decline in metacognitive performance as these studies, but a different pattern regarding the modulation of the $P_{e/c}$ amplitude by ageing, implies that the $P_{e/c}$ does not seem to reflect the individual degree of metacognitive accuracy, but rather a subjective feeling of confidence in errors that may be more or less strongly related to objective accuracy. Again, the different findings regarding the modulation of the $P_{e/c}$ amplitude by age might be related to the employed type of metacognitive assessment (i.e., binary error detection vs. multi-point confidence) because the processing of neural signals to overt ratings requires additional steps (e.g., the transformation of a sense of confidence into a binary judgement) that might be susceptible to noise and differ between individuals (Desender, Boldt, et al., 2019). Ultimately, however, it is more relevant to relate confidence-related neural signals to the adaptation of behaviour rather than to a certain type of metacognitive judgements if we are interested in their functional relevance. This has been done by Desender and colleagues who showed that single-trial variation in the $P_{e/c}$ amplitude predicted response caution in the subsequent trial (Desender, Boldt, et al., 2019) and choices to seek or not seek more information before committing a response (Desender, Murphy, et al., 2019). It would be important for future research to test whether this function of the $P_{e/c}$ is affected by ageing, independent of its direct relation to metacognitive reports.

4.1.2.4. Limitations

In this study, EEG analyses focussed on two well-defined ERP components that have frequently been assessed in error monitoring studies. Accordingly, we did not consider other ERP components or oscillatory brain activity, which are also known to be related to error processing (Cavanagh & Frank, 2014). As mentioned in the introduction (1.3.2.5), the CPP/P300 is discussed as another key candidate to index graded levels of decision confidence (Gherman & Philiastides, 2015; Parés-Pujolràs et al., 2020). Rausch and colleagues (2020) investigated the time course of confidence formation by using a novel approach of combining EEG analysis and cognitive modelling to identify correlates of confidence. In a masked orientation task, they manipulated the stimulus strength and argued that a neural correlate of decision confidence should mirror the statistical regularities of confidence judgements as being modulated by stimulus strength. Examining three candidate ERP components, namely the stimulus-locked CPP/P300, and the response-locked $N_{e/c}$ and $P_{e/c}$, it was shown that the modulation of the CPP/P300 by stimulus strength was resembling the modulation of confidence by stimulus strength across correct and incorrect responses most.

Notably, however, Feuerriegel and colleagues (2022) revealed widespread methodological issues in the measurement of ERP components that question previous findings. Regarding the $P_{e/c}$ it was shown that its modulation by confidence critically depended on the choice of the baseline. Using a *pre-response* baseline to compute the P_e, as is typically done (Boldt & Yeung, 2015; Desender, Murphy, et al., 2019; Rausch et al., 2020) and as we did, the authors argued that its amplitude would be confounded by the pre-response CPP/P300 component, which scales positively with confidence. When the resulting CPP/P300 difference in the pre-response baseline is corrected to zero, this inflates the negative relationship between P_e amplitude and confidence following the response. Regarding the CPP/P300, the authors provided evidence that this component might in fact reflect overlapping activity from a slightly later and more frontal ERP component. Together, these findings illustrate that no agreement regarding the neural correlates of confidence has yet been reached. While this clearly impedes the assessment of age-related changes in these processes, the study of ageing might also contribute to advancing our understanding of neural processes underlying confidence judgements. For example, our observation of a lack of a modulation of the P_e by age might suggest that the $P_{e/c}$ indexes a subjective sense of confidence that is independent of the person's metacognitive abilities.

4.1.2.5. Conclusion

In sum, we concluded from our findings in Study 1b that the study of the neural correlates of error monitoring using confidence judgements offers valuable insights both into the relationship between these two research fields and the functional role of the ERP components in metacognition. While our results did not replicate a relationship between age-related impairments in metacognitive performance and the $P_{e/c}$ amplitude, we suggest that a combination of multiple sources might contribute to this specific deficit, one of them being age-related alterations in the $N_{e/c}$ amplitude. This proposal will be discussed in more detail in the following Section 4.2.1.

4.1.3. Study 2

In Study 2, we aimed to further characterise age-related changes in metacognitive performance by quantifying the relationship between judgements of decision confidence and the behavioural parameters of the action reporting the respective decision. The study was based on the same experiment as Study 1, using the same behavioural data (RT, accuracy) and additionally recordings of the response force. Force was recorded using custom-made force sensitive keys, which has a twofold advantage of having a much higher temporal resolution than standard keyboards used to record responses (Shimizu, 2002) and having higher sensitivity to detect micro-movements like twitches, partial responses, or error corrections that would remain unnoticed when only registering the first time point when a certain force threshold is crossed, as in standard response devices. Such micro-movements might affect the timing or force of the activity registered as the response (e.g., the 'actual' response will have a longer RT when it follows an ipsilateral or contralateral partial response; Gajdos et al., 2019). Additionally, they might distort confidence judgements, if, for example, the participant mistakenly relates the judgement to a correction response while the initial premature movement has incorrectly been registered as the response.

4.1.3.1. Relevance of response force

Of main interest for us, however, was the produced force of responses, of which we derived the maximum, that is, the peak force (PF). We used PF because it indexes cognitive processes that can be dissociated from processes associated with RT as they have been shown to be functionally independent (Cohen & van Gaal, 2014; Franz & Miller, 2002; Jaśkowski et al., 2000; Stahl & Rammsayer, 2005). For example, it was shown that a task-irrelevant stimulus increased either response speed or force, depending on when it was presented in temporal relation to the imperative stimulus (Stahl & Rammsayer, 2005). Our findings of a differential association between confidence and the response parameters RT and PF further support the notion of functional independence between both response parameters. Previous research has identified several stimulus-related characteristics that increase PF, like increased conflict (Kantowitz, 1973; Van Der Lubbe et al., 2002), temporal stimulus uncertainty (S. Mattes & Ulrich, 1997), and low stimulus probability (S. Mattes et al., 2002). Moreover, it was widely reported that PF is positively related to the strength of sensory stimuli, that is, the more sensory evidence was available, the more force was exerted to report a decision (Jaśkowski et al., 2000; Ulrich et al., 1998). Given the close relationship between sensory evidence strength and confidence in a decision, a relationship between PF and confidence seems likely.

In addition, research on the computation of confidence judgements has revealed a crucial role of motor activity in the evaluation process. For example, (1) disruption of the motor system reduced metacognitive accuracy in a perceptual discrimination task (Fleming et al., 2015), (2) metacognitive accuracy was disrupted by manipulating the movement speed of the initial response (Palser et al., 2018), (3) confidence and metacognitive accuracy were higher when the decision required a motor response compared to when it was indicated by not moving (Siedlecka et al., 2021), and (4) confidence was higher when higher physical effort had to be exerted to report a decision (Turner, Angdias, et al., 2021). Given this close relationship between motor activity (and PF in particular) and metacognition, and the described differences in metacognitive accuracy across the lifespan, we are compelled by the association between PF and confidence to reason that it might vary as people age, too.

4.1.3.2. Effects of ageing on response parameters of confidence

In our experiment, correct responses were given faster by younger adults when the subsequent confidence judgement was higher and errors were reported with higher force when the subsequent confidence judgement was lower, indicating that the error was (likely to be) detected. Indeed, these relationships changed across the lifespan. The effect for correct responses was pronounced with older age, while the effect for errors, on the other hand, was only found for younger adults until 44 years of age. This study comprises the first systematic assessment of the relationship between confidence, response parameters, and age.

Study 2 was of clear exploratory nature regarding the direction of a potential age effect, and the primary aim was to quantify the relationship between confidence and action-related parameters. Consequently, the correlational results cannot and were not intended to be used to derive causal explanations of the different formation of confidence across age. However, in the manuscript, we offered two alternative (and possibly complementary) interpretations of our findings: first, the subjective sense of confidence might drive the response parameters or they might jointly be informed by a third variable (Kiani & Shadlen, 2009; Vickers, 1979). This means, for example, that strong evidence for one response option in an easy task might lead to high confidence, which again might lead to a faster and more forceful response. Alternatively, response-related information might be one of various sources contributing to the formation of confidence. As such, also information that is not available for the initial decision (like RT or PF) might be used as a cue about the ease of the decision process (Gajdos et al., 2019; Kiani et al., 2014; Turner, Angdias, et al., 2021). This interpretation will be discussed in more detail in the following section.

4.1.3.3. Modulation of confidence by response parameters

The suggestion that confidence may be modulated by response parameters is consistent with the assumptions of the model by Fleming and Daw (2017), suggesting that the monitoring system should make use of all available information in order to improve the accuracy of metacognitive judgements, one of which might be information about the motor response. This may be particularly constructive when sensory evidence is limited or ambiguous and does not provide sufficient information about the decision accuracy. The strong decline of metacognitive performance with increasing age could suggest that some sources of input to the monitoring system were missing or could not be accessed by older adults. This might apply to sensory evidence, but as we observed metacognitive accuracy to decline beyond task performance and other studies showed an age-related decline in metacognitive performance for comparable task performance (Niessen et al., 2017; Palmer et al., 2014) response-related information that was not available for the initial decision but contributed to confidence might have additionally been restricted in older adults.

For correct responses, we replicated a widely reported negative relationship between confidence and RT, that is, higher confidence was related to faster responses (Fleming et al., 2010; Rahnev et al., 2020). In our study, the negative relationship was pronounced in older adults, which was likely due to their larger RT variability in correct responses compared to younger adults. Assuming a causal effect of RT on confidence, however, one might also speculate how this finding might contribute (a small fraction) to the age-related decline in metacognitive accuracy. It is possible that the learnt association between confidence and RT was more relevant for older compared to younger adults as a cue to inform their metacognitive evaluation because their ongoing accumulation of sensory evidence might yield little or ambiguous information about the decision accuracy. Hence, older adults might have over-relied on the confidence-related information conveyed by the RT, being less able to flexibly adjust the weight they assign to this information (e.g., a participant might be distracted for a moment, but then re-orient their attention and give a correct response in an easy decision; thus, RT would be long, but confidence should nevertheless be high). This might then lead to the lower resolution of confidence ratings of older adults in correct responses. For further discussion of the age-related changes in the relationship between confidence and RT, I refer to the manuscript of Study 2 (3.2.2) and will now elaborate more on the relationship between confidence, PF, and age, which constitutes a novel aspect of our study.

We found a significant relationship between confidence and PF for errors, which vanished with older age. While PF is investigated considerably less often than RT, two recent studies directly investigated the relationship between PF and explicit metacognitive judgements of error commission or confidence. Turner, Angdias, and colleagues (2021) used a perceptual discrimination task and a full range confidence scale similar to ours. They instructed participants (after the response was initiated) to produce different degrees of PF in order to report a decision and it was shown that higher confidence could be predicted from higher levels of physical effort. Intriguingly, we could also show an association between PF and confidence in younger adults despite two marked differences to the study by Turner, Angdias, and colleagues: first, participants in our study did not pay attention to their force production nor received feedback about the exerted force, while in their study, participants were instructed and trained to produce certain levels of force and received visual feedback about the exerted force. Moreover, due to the different recording devices, the overall level of force was much lower in our study (40 cN response threshold vs. 20-60% of maximum grip force in Turner, Angdias, et al., 2021). While the Turner, Angdias, and colleagues' (2021) study found a positive association between PF and confidence, we only found a significant association for errors, which was reversed in direction, that is, higher PF was related to *lower* confidence in Study 2. Instead, our findings are in line with a recent study using the same recording device as ours and a complex version of the Simon task (Stahl et al., 2020). In this study, detected errors were reported with higher force than undetected errors, which mirrors our findings of higher PF in errors that were rated with low confidence. One possibility to jointly explain the seemingly contradicting results of these three studies is the assumption of an association between PF and *certainty* instead of confidence. PF was high when certainty in having made an error was high (Stahl et al., 2020 and Study 2) and high PF boosted participants' certainty when they were correct (beyond the primary effect of sensory evidence on confidence; Turner, Angdias, et al., 2021). The latter study also found PF to boost confidence in errors, however, they used a perceptual discrimination task in which participants are rarely very certain they made an error. Thus, the monitoring system might interpret high PF as indexing high sensory evidence for the first-order decision (Jaśkowski et al., 2000; Ulrich et al., 1998), which in turn might contribute to the computation of confidence by increasing high confidence in correct responses further and decreasing low confidence in errors further (i.e., increasing certainty in having made an error). In fact, studies on error monitoring suggest that errors are more likely to be detected when stronger evidence is available (Ullsperger et al., 2010).

4.1.3.4. A proposed mechanism for the effect of RT and PF on metacognitive processes

The results of this study are consistent with the view that decisional and metacognitive processes are closely interconnected and operate via multiple, reciprocal feedback cycles (Siedlecka et al., 2021). I propose that the monitoring system might accumulate decision-relevant information in an abstract way, for example, by evaluating the amount of evidence that the decision is based on. In such a framework, RT might reflect the ease of a decision (Susser & Mulligan, 2015). Very fluent processing would lead to a fast response, and at the same time, the clear sensory evidence and additionally the proprioceptive feedback about the ease of the response (i.e., the RT) will be used to form a sense of confidence. PF, on the other hand, might reflect the amount of available evidence for a decision. As such, conflicting evidence (e.g., similar colours to choose between) might nevertheless lead to a forceful response if a lot of evidence is available (e.g., presented for a long time), and the resulting high PF of the initial response might increase the certainty of the metacognitive judgement. In other words, the post-decisional accumulation of conflicting evidence may lead to the confirmation of the initial decision or to a change of mind, and the information about the PF of the response might refine the confidence judgement by increasing the certainty in having made a correct or incorrect decision in case of a high amount of evidence, and decrease the certainty in case of a low amount of evidence.

To illustrate, these suggested mechanisms might operate in a serial order (Fleming & Daw, 2017): first, the monitoring system might evaluate the ease of a decision as reflected in the RT. If this is short, the simultaneously arising confidence will be boosted. Instead, if the RT exceeds a certain limit, the monitoring system might try to resolve the conflict and in a second step, evaluate the amount of available evidence as reflected in the PF to refine the outcome of the metacognitive evaluation process.[2]

As suggested by Filevich and colleagues (2020), such monitoring processes might be flexibly adjusted, and can vary inter- and intraindividually with regards to the choice and emphasis

---

[2] Fleming and Daw (2017) explicitly clarify that second-order models of metacognition do not restrict the computation of confidence to occur in a serial manner, but that it can also be accommodated by a parallel architecture, which can explain for example very fast error detection and correction.

of different sources of information that are incorporated into confidence judgements. This could also account for age-related changes in the relationship between confidence and response parameters. Our results, for example, could be explained as a reduced trust of older adults in their fine motor skills and the related informative value of PF about the decision accuracy (Rosi et al., 2018). Accordingly, high PF would not mean certainty is boosted, so the lack of relationship between confidence and PF in older adults could be observed. However, control analyses of our data revealed that, in contrast to previous findings (Sosnoff & Newell, 2006), the force output of older adults was not more variable compared to younger adults, which might question the plausibility of this interpretation. In general, growing evidence highlights a relevant role of response force in metacognitive processing, but in contrast to RT, more studies are needed to define how exactly PF is modulated as well as its potential function in the computation of confidence.

4.1.3.5. Limitations

The use of custom-made force keys could serve as an advantage but also a disadvantage as it makes it difficult to directly place our findings beside previous studies which mostly used devices that required and recorded markedly higher levels of force (e.g., Hagura et al., 2017; Jaśkowski et al., 2000; Nguyen et al., 2021; Turner, Angdias, et al., 2021). Despite this, our keys have successfully been applied in multiple studies from the lab (e.g., Armbrecht et al., 2012, 2013; Siswandari et al., 2019; Stahl et al., 2015, 2020) and are especially informative because they comprise a range of force that is relevant in many real life situations, like typing on a standard keyboard or using touch pads (Y. T. Ko et al., 2012).

An open question that follows from this study is whether the proposal of PF tracking the certainty of a decision, independent of the direction of the metacognitive judgement, proves true. This could be tested, for example, by using a two-step metacognitive judgement (i.e., first choosing between ‘error’ and ‘correct’ and afterwards indicating the certainty in the prior choice), and computing the relationship between PF and the certainty judgement.

Finally, this study employed a correlational design, for which I stated the reasons and benefits before. This is not an issue in itself, but it limits the interpretability of our findings with regards to the mechanisms underlying the age-related decline in metacognitive performance. Having established an association between confidence and naturally occurring response parameters, future studies could manipulate RT and/or PF, for example by providing visual feedback about the current and the required degrees of speed/force, or by instructing participants to response faster/slower or stronger/weaker than they naturally would (Palser et al., 2018; Turner, Angdias, et al., 2021). However, both alternative also come with downsides (e.g., reducing the focus on proprioceptive feedback, affecting primary task performance) and might alter the relationship to confidence judgements. If, in such an experiment, our observed pronounced negative relationship between RT and confidence in older adults persists, it could be revealed whether this has a positive or a negative effect on metacognitive accuracy, thus having a compensatory or a detrimental function.

4.1.3.6. Conclusion

In conclusion, we showed in Study 2 that confidence was related to fine-grained changes in response parameters of the first-order decision. The differential association of confidence in correct responses and errors with RT and PF, respectively, supports the notion that RT and PF might index distinct cognitive processes. Notably, the observed relationships changed with age. This might reflect distorted proprioceptive feedback about response dynamics, an inability to integrate action-related information into processes like the computation of confidence, or an inappropriate interpretation of the information carried by the response parameters. The question whether the observed changes in these relationships have an effect on metacognitive accuracy could not be resolved in this study and offers a potential avenue for exploration.

4.2. Implications of research findings

In this final section, I will discuss the implications of the findings presented in this thesis. For this, I will bring the findings of the three studies together and evaluate them in the context of recent evidence from the literature and established theories of neurocognitive ageing. First, I will tackle the fundamental goal of understanding the age-related decline in metacognitive performance from different perspectives. Then, I will take a step back and consider the functional relevance of the derived conclusions for everyday life and which avenues for interventions they might hint at.

4.2.1. Why does metacognitive performance decline with age?

4.2.1.1. Characteristic decline of metacognitive functions with older age

In the studies described in this thesis, we found a marked decline in metacognitive accuracy with older age. As metacognitive functioning scales with age-related changes in prefrontal brain networks across the lifespan, this general pattern can be expected from a developmental perspective on metacognition. Looking at the development of metacognitive performance from childhood to late adulthood, it generally increases during adolescence, plateaus in adulthood, and declines with older age (Hämmerer et al., 2014; Weil et al., 2013). In adolescence, relevant brain regions and their functional connections are still maturing, while in older age, as predicted by frontal lobe theories of ageing (West, 1996), they are particularly prone to a reduction of white and grey matter, which directly affects their functioning (Weil et al., 2013).

As executive functions also rely on prefrontal brain networks, it could be expected that an age-related decline in executive functions can explain the decline in metacognition, given their close conceptual similarity (Filippi et al., 2020). While we did not directly measure executive functioning in our study, Palmer and colleagues (2014) did so. They investigated metacognitive accuracy in a similar sample as the one presented in this thesis in a visual perception task and also revealed a strong decline in metacognitive accuracy. Moreover, this study assessed executive functioning using a neuropsychological test of the ability to shift attention between two tasks (Trail Making Test; Reitan, 1992) and showed that the score of this test did not predict metacognitive efficiency across participants. While it is difficult to entirely separate the two constructs or to exclude an indirect effect of age-related limitations in executive functioning on metacognitive performance, it seems likely that our results reflect an actual impairment in metacognitive monitoring in older adults that cannot simply be explained by age-related declines in executive functions.

4.2.1.2. Increased uncertainty in older adults

Importantly, Palmer and colleagues (2014) measured confidence on a scale from guessing to certainty in being correct, while we allowed to indicate certainty in having made an error and thus, bridged the gap to studies using error detection judgements. In fact, much of what we know about effects of age on metacognition in decision-making has come from the field of error monitoring. More than 30 years ago, Rabbitt (1990) reported a selective decline of older adults to signal errors in a choice reaction time task. This finding has been replicated in many subsequent studies using different perceptual tasks and manipulations (Endrass, Schreiber, et al., 2012; Harty et al., 2013, 2017; Niessen et al., 2017; Wessel et al., 2018). Together, confidence studies using a scale with guessing as an endpoint or studies using binary forced choice or error signalling responses might miss meaningful differences within the low confidence range or within correct trials, respectively. Reports of decreased error detection rates in older adults may thus be interpreted as a general overconfidence in older adults, but the assessment of confidence on a multi-point scale allowed us to unravel older adults' metacognitive rating behaviour in more detail. We showed by considering errors and correct responses that older adults generally tended to give more conservative confidence ratings, that is, in correct responses, they were *not* overconfident. This observation of a higher ratio of uncertain confidence ratings in older adults might have two reasons: first, the ratings might indicate an intentional or unintentional bias in criterion setting, or they might indicate actual uncertainty in the evaluation of decision accuracy.

In support of the former interpretation of age-related biases in criterion setting, McWilliams and colleagues (2022) also observed more uncertain confidence ratings with older age. In their study, age showed a negative effect on confidence *bias* which was strongly correlated within individuals across the two studied domains. This is in line with previous findings of a strong intraindividual, cross-domain generality of confidence bias (Ais et al., 2016; Lehmann et al., 2022). Additionally, in their large online study McWilliams and colleagues (2022) asked participants to estimate how well they would perform and how well they performed compared to others before and after completing the tasks, respectively. The obtained reports were closely related to metacognitive bias showing lower mean confidence with older age. This bias towards more conservative confidence ratings suggests an age-related shift in criterion setting, that is, the internal threshold to report high certainty was very high for correct responses and very low for errors. Based on evidence that decision criteria can intentionally be controlled (Steinhauser & Yeung, 2010), it could be speculated that his behaviour reflects a response strategy that is intended to show modesty (McWilliams et al., 2022). Alternatively, it might also reflect a generally reduced self-efficacy in older adults that may lead to high doubt about the own metacognitive capacities. In fact, trust in the own cognitive abilities was shown to decline from an age of 50 years (Rosi et al., 2018).

In support of the latter interpretation that confidence ratings indicate actual uncertainty, older adults might in fact have similar criteria as younger adults, but rarely accumulated enough evidence to cross the internal threshold for high certainty, which would lead to the more conservative confidence judgements in older adults. In other words, the computation of confidence may be noisier in older adults, leading to higher uncertainty about most decisions.

In the following, I will discuss how the results of Study 1b and 2, showing altered behavioural and neural correlates of confidence, support this assumption. In Study 1b, we aimed to identify age-related changes in the neural processing of decision confidence that might underlie the observed decline in metacognitive performance that accompanies ageing. We studied the $P_{e/c}$, an established marker of error detection (Nieuwenhuis et al., 2001) that has also been shown to scale with reported decision confidence (Boldt & Yeung, 2015; Scheffers & Coles, 2000). Previous studies have reported a decline in the P_e amplitude of detected errors in older adults that reflected an attenuated error detection rate (Harty et al., 2017; Niessen et al., 2017). Our findings contradict these studies as we observed an age-equivalent modulation of the P_e by confidence in a way that P_e amplitudes were higher for errors rated as ‘surely wrong’ compared to errors rated as ‘surely correct’. These findings may be explained in terms of an evidence accumulation process. According to an overarching framework proposed by Desender and colleagues (2021), the computation of metacognitive judgements can be explained, similar to other types of decisions, as the post-decisional accumulation of evidence over time (Ratcliff et al., 2016; Shadlen & Kiani, 2013). Evidence is accumulated until it reaches a certain decision threshold that may reflect, for instance, error detection or a certain degree of confidence. It has been suggested that the $P_{e/c}$ might index the momentary accumulated evidence in favour of an error after the response that reaches a stereotyped amplitude whenever an error is signalled or a low confidence rating is given (Desender, Ridderinkhof, et al., 2021; Murphy et al., 2015). Accordingly, our findings might indicate that the accumulation of post-decisional error evidence was intact up until older age, and that older adults in fact set similar criteria as younger adults for choosing a certain confidence level because the single-trial P_e amplitudes for a given confidence level were comparable between younger and older adults. Hence, as older adults had a lower rate of low confidence errors (i.e., errors that were likely detected) they may have accumulated enough post-decisional evidence to reach the threshold for error detection in significantly fewer cases. This explanation could account for our observed stable $P_{e/c}$ amplitudes across age co-occurring with a decline in metacognitive accuracy.

Divergent findings compared to error monitoring studies might thus be related to the employed metacognitive assessment. Since error detection studies only allow for a binary rating of a decision being correct or incorrect, participants may report an error despite a certain degree of uncertainty, which they could have rated as ‘maybe wrong’ on a multi-point rating scale. This would result in an age-related decline of P_e amplitudes of errors that were signalled as errors because they would include a considerable number of trials with higher uncertainty, which is related to lower P_e amplitudes. Taken together, the use of a metacognitive assessment that allowed to indicate uncertainty in our study and the modelling of confidence in relation to single-trial variation in the P_e amplitude showed that impaired error evidence accumulation does not appear to cause age-related declines in metacognitive accuracy as it was reported in error monitoring studies (Harty et al., 2017; Niessen et al., 2017).

To give an interim summary, I have presented evidence for higher uncertainty in confidence ratings of older adults. Our findings together with previous work suggest that error evidence is similarly accumulated across age, implying that increased uncertainty might have another origin. So, what else could make older adults doubt their own decisions to a degree that surpasses their decline in primary task performance? Here, I will propose that one aspect that is contributing to age-related declines in metacognitive performance may be the process indexed by changes in the amplitude of the $N_{e/c}$. We found that the N_e but not the N_c declined with age, which was driven, in particular, by a decline of the N_e amplitude of errors that were rated with a low confidence (i.e., likely detected as erroneous). This resulted in a smaller difference between the N_e amplitudes of different confidence levels in errors for older adults.

4.2.1.3. Attenuated representation of the correct response in older adults

As outlined in Chapter 1, Section 1.3.2.3, different functional theories of the $N_{e/c}$ have been proposed. Here, I will briefly explain each theory and discuss how our findings might contribute to explaining age-related deficits in metacognition within these frameworks. According to mismatch theories, the $N_{e/c}$ reflects conflict between the (efference copy of the) executed response (i.e., the erroneous response) and the best estimated representation of the correct response that emerges due to continued processing of the stimulus (Coles et al., 2001; Falkenstein et al., 1991). Hence, an error is conceptualised as a premature response due to incomplete stimulus processing. Conflict monitoring theories similarly predict that the $N_{e/c}$ reflects conflict, but more specifically conflict between different simultaneously activated response tendencies (Botvinick et al., 1999). According to this view, the $N_{e/c}$ is larger in errors compared to correct responses because the cognitive processes of the tendency that led to the incorrect response and the tendency towards the correct response that emerges during ongoing evidence accumulation interfere (Yeung et al., 2004). Thus, the $N_{e/c}$ is proposed to reflect more general post-response conflict that is not restricted to (the detection of) errors, but can also occur in correct trials with high conflict. Lastly, reinforcement learning theory postulates that errors violate the prediction of a correct response and are thus events that are worse than expected (Holroyd & Coles, 2002). These prediction errors trigger a learning signal that is propagated from subcortical areas to the neural generators of the $N_{e/c}$, and is reflected in its amplitude. Translated to our findings, in case of low confidence errors, older adults might have experienced less conflict than younger adults, either because of an attenuated representation of the executed or the correct response, similar post-response correct and incorrect response tendencies, or because errors were generally less unexpected or surprising in older adults given their higher error rates.

Taken together, all of these theories require a representation of the correct response, or at least some evidence in favour of the correct response to be present, in order to elicit a $N_{e/c}$. This functional role of the $N_{e/c}$ was experimentally confirmed by Di Gregorio and colleagues (2018). In this study, the authors introduced one condition in which no target was present and thus, the correct response could not be represented internally. However, a metacognitive evaluation of the response was still possible because task-irrelevant stimuli (flankers) were always associated with an incorrect response and thus, responding to the flankers could be detected as an error without seeing the target. Indeed, these trials did not elicit a $N_{e/c}$. Further, the suggestion that a representation of the correct response may be distorted in older adults finds support in studies investigating the effect of response-set size on error monitoring (Maier et al., 2010; Rabbitt, 1967). Increasing the response-set size impeded response selection and was related to smaller differences between N_e and N_c amplitudes because the post-response information processing was less likely to result in a correct response tendency (reducing N_e amplitudes) and more likely to lead to an erroneous tendency (larger N_c

amplitude; Maier et al., 2010). Accordingly, if older adults in our study had greater difficulties to remember the stimulus-response mapping throughout the experiment, response selection might have been complicated, resulting in weaker representations of the correct response after continued processing (Stahl et al., 2020). Thus, if older adults in our study were faced with weaker top-down error signals that could enter the metacognitive process of performance evaluation, this might have restrained the accuracy of later metacognitive judgements because they were missing the reference frame for further computation, namely, whether later, independent error evidence signals (accumulating evidence from different or additional types of information unavailable for early error monitoring; reflected at the neural level in the P_e amplitude; Charles et al., 2013; Di Gregorio et al., 2018; Wessel et al., 2011) were to be interpreted in the context of high or low conflict.

Theories of structural and functional neural changes in normal ageing postulate that these alterations in neural activity might either reflect compensatory mechanisms as recruiting additional resources to be able to achieve comparable task performance to younger adults, or alternatively, reflect an impaired mechanism of recruiting specialised brain regions for a given task (Cabeza, 2001; Heuninckx et al., 2008). These theories have recently also been tested using the EEG and functional connectivity analyses (Rosjat et al., 2020). This study provided evidence for overall increased but less flexible connectivity in older adults that might make the recruitment of task-specific brain regions more difficult, supporting the dedifferentiation hypothesis. In our study, we focussed our electrophysiological analyses on two central electrodes and can thus not inform the debate between the two theories of spatial activation changes accompanying ageing. Nevertheless, at these two regions of interest, we did not find evidence in support of the compensation view in terms of neural activity since older adults did not show overall increased ERP amplitudes. Rather, the age-related changes that we did observe in the electrophysiological correlates of error monitoring point towards a decreased specificity of neural processes related to metacognition in older age. That is, neural activity related to the $N_{e/c}$ was less confidence-specific in errors committed by older compared to younger adults, possibly contributing to less efficient metacognitive processing with older age. This may be caused by age-related atrophy in the prefrontal brain regions supporting metacognitive monitoring (Dully et al., 2018).

4.2.1.4. Multiple sources contribute to age-related metacognitive decline

So far, I have argued that processes that are reflected in age-related changes in the $N_{e/c}$ amplitude in high conflict situations may be one aspect that is contributing to less accurate confidence judgements in older adults. However, alternative or additional factors are conceivable. Some models, such as the one proposed by Fleming and Daw (2017), suggest that multiple sources of input may inform post-decisional error evidence accumulation processes. This follows because confidence was related to processes that are not directly relevant for the task and did not affect task performance, such as sensory uncertainty (Charles et al., 2013), unexpected arousal (Allen et al., 2016), or changes in autonomous nervous system activity (Wessel et al., 2011). Building on this assumption, a number of recent studies have shown a causal relationship between motor activity of the response and confidence (Fleming et al., 2015; Palser et al., 2018; Siedlecka et al., 2021; Turner, Angdias, et al., 2021). It was suggested that the motor response might provide the metacognitive evaluation with additional information about the performance of the decision system. For example, Siedlecka and colleagues (2020) found that the accuracy and the mean of confidence ratings were higher when given after actively reporting the decision compared to a passive condition. They proposed that the motor response might serve as a cue that the decision has successfully been carried out, boosting decision confidence. Another interpretation of the positive effect of motor-related information in metacognitive accuracy is that learnt associations between response parameters and the ease of the decision are used to inform confidence judgements (Gajdos et al., 2019; Kiani et al., 2014).

While we did not set out to directly test this possibility, we found evidence in support of a relationship between confidence and motor-related information in Study 2 by showing that confidence was related to at least two behavioural parameters of the response, namely the time and the force taken to report a decision. Notably, we found these relationships to change with age. One possible explanation of these results is that, if response parameters contribute to sharpening confidence judgements in younger adults, this learnt relationship might be impaired in older adults. According to Fleming and Daw's model, this is possible since the decision and confidence are computed separately and the action-related information that informs confidence in younger adults is not accessible for the primary decision. However, we have discussed alternative interpretations of the results of Study 2 in the previous Section 4.1.3.

The results of the studies presented in this thesis, together with findings from the literature, suggest that a mixture of neurophysiological and behavioural changes contribute to the decline in metacognitive performance associated with healthy ageing. Specifically, I pointed out that the decline might often be the result of more conservative confidence ratings in older adults. While these may be related to confidence bias, I discussed how our results rather support the interpretation of increased uncertainty as arising from distorted early error signals reflected in the $N_{e/c}$ amplitudes as well as restricted access to certain (motor) sources of information during the computation of confidence.

An interesting next step for future research would be to directly test whether the age-related decline in metacognitive accuracy reflects a strategic choice or an actual deficit. Steinhauser and Yeung (2010) provided evidence in young participants that it is possible to shift internal decision criteria for reporting an error via manipulation. This could also be applied to confidence ratings, for example by using a cover story that incites a liberal or conservative response strategy (e.g., by advocating for modesty), or by rewarding different decision policies (e.g., rating an error as erroneous with high certainty will yield a larger reward than for rating it as erroneous with low certainty). If older adults would achieve higher metacognitive accuracy when prompted to give more liberal confidence judgements, this would suggest that their apparent metacognitive impairment is rather a voluntary behaviour.

Besides, future work could continue collecting evidence to answer the ultimate question of how the monitoring system decides which information is integrated to what extent into metacognitive judgements. A first step to answer this question could be, for example, to separately manipulate different of the suggested sources of evidence (e.g., information about conflict, as reflected in $N_{e/c}$ amplitudes, using a masking paradigm or instructions intended to alter response parameters; (Di Gregorio et al., 2018; Palser et al., 2018; Rausch et al., 2020; Turner, Angdias, et al., 2021) and to assess the degree to which the combination of manipulations affects metacognitive performance as a function of age.

4.2.2. What are practical implications of age-related changes in metacognitive accuracy?

In the end, the question remains why metacognition is relevant at all. What are the consequences associated with high or low metacognitive accuracy and how does it have an impact on everyday life, especially in older adults? While metacognitive evaluation may have an intrinsic, self-reflective value, the ultimate goal is to improve future behaviour. Consider a decision maker who recognises a number of errors in a row, they should invest some effort into deriving better decision outcomes. As this example illustrates, we typically assume a link to conscious awareness, a phenomenal experience related to the just made decision. Indeed, reporting metacognition as a confidence rating or an error detection response requires conscious access to the metacognitive evaluation. However, a substantial literature has reported that a number of monitoring processes operate automatically and thus subconsciously (Fleming, Dolan, et al., 2012; Ullsperger et al., 2014; van Gaal et al., 2012). For example, it was shown that errors can be rapidly corrected without participants becoming aware of it. While participants normally detect the majority of their errors in experiments, a substantial number of correct responses are preceded by a partial error, that is, a subliminal activation of the incorrect response effector. Notably, even though partial errors and their correction are rarely consciously perceived, they are nevertheless immediately and successfully corrected (Ficarella et al., 2019; Gajdos et al., 2019). Furthermore, participants were found to slow down after errors even if they were not conscious due to stimulus masking (Cohen et al., 2009). From these finding, it was concluded that a basic metacognitive evaluation does not require consciousness to have a positive effect on subsequent behaviour. However, the likelihood of error detection is obviously significantly larger, when the errors are consciously perceived (Charles et al., 2013; Cohen et al., 2009; Niessen et al., 2017).

In this thesis, I focussed on metacognitive processes that were reportable in confidence judgments and hence, supposedly conscious. In Study 1a, we investigated the effect of confidence judgements on decision policies in the following trial. Interestingly, we showed that when people age, their decline in metacognitive performance does not translate to the short-term adaptation of behaviour. At a single-trial level, participants of all ages responded more cautiously after decisions which they doubted or after which they changed their minds. This means that the decisions were on average slower, but also more likely to be correct. In these trials, the subjective feeling of low confidence might have prompted participants to seek more information in the next trial in order to make a better decision with higher confidence (Desender et al., 2018). This finding points towards a specific aspect of metacognition that seems to be well preserved until old age. Still, we also showed that confidence judgements of older adults matched the objective performance to a smaller degree than those of younger adults. Hence, older adults successfully adapted their behaviour depending on the internal evaluation of their decision (i.e., confidence), which was their best available estimate as no objective feedback was provided. However, if the metacognitive evaluation has a low resolution, the behaviour can rarely be adapted when required. In other words, when the decision is wrong but confidence is high, the behaviour will not be adapted and vice versa, when the decision is correct but confidence is low, the participant will probably slow down even though this would not have been needed. Thus, if the metacognitive judgement itself is not accurate, the subsequent adjustment of behaviour that results from the evaluation may not be the accurate choice in this case. This might then result in more errors when response speed is not reduced after undetected errors and an additional slowing if low confidence leads to higher response caution despite a correct decision. In short, our results point towards a dilemma regarding the metacognitive performance of older adults: despite revealing an aspect of metacognitive processing that seems to be stable across age (i.e., the behavioural adjustment in relation the *best available estimate* of the decision accuracy), the low resolution of these estimates causes the adjustments to rarely be appropriate.

The assumption that older adults will adjust their behaviour less effectively due to the lower resolution of confidence ratings is further supported by recent evidence linking variations in $N_{e/c}$ amplitude to the adjustment of subsequent trial decision threshold and drift rate (in terms of signal detection theory; indicating, respectively, a higher internal criterion to commit to a decision and a stronger focus on task demands; A. Mattes et al., 2022). Across three experiments, it was shown that larger $N_{e/c}$ amplitudes predicted slower responses and higher accuracy at a single-trial level. While we did not assess the relationship between post-response behaviour and ERP amplitudes in Study 1b, we found reduced N_e amplitudes of low confidence errors with increasing age. This suggests that older adults' capacity to adjust their behaviour to current demands may also be related to their reduced sensitivity to conflict, as it might be reflected in the attenuated N_e amplitudes (see Section 4.2.1).

Hence, the following question is whether this self-reinforcing age-related impairment in adjusting behaviour adaptively (i.e., response caution is aligned to confidence, but low metacognitive accuracy leads to maladaptive adjustments, which again might lead to more errors and additional slowing of responses) can be counteracted. Indeed, evidence suggests that metacognitive training can be effective to a certain extend. First of all, simply engaging in a task that requires metacognitive evaluation was already found to improve performance (Bonder & Gopher, 2019). It was shown that a group of participants that indicated their confidence after a perceptual decision performed significantly better and continued improving in the primary perceptual task compared to a control group without confidence rating. This seems like a promising endeavour for older adults because avoiding errors in the first place reduces the need for metacognitive monitoring and control. However, it also seems possible to train the metacognitive monitoring itself. A very recent study reviewed evidence for the effect of training on metacognitive efficiency (Katya & Fleming, 2022). While previous studies were inconclusive regarding positive effects of training, this meta-analysis used a Bayesian modelling approach to estimate the effect of training on session-by-session changes. They found that metacognitive efficiency in a test session without feedback was slightly, but significantly improved after receiving a number of adaptive training blocks where feedback about the resolution of the confidence judgements paired with a monetary reward was provided. Notably, the training effect was enhanced for individuals with initially lower metacognitive efficiency (Katya & Fleming, 2022), which might make this specific training protocol suitable to apply in the training of older adults. Moreover, *when* positive effects of training on metacognitive performance were found, they generalised from the trained perceptual domain to the memory domain (Carpenter et al., 2019).

Lastly, it can be asked how generalizable the findings of behavioural adaptations as well as training effects on metacognitive performance are to real life situations where metacognition is not assessed in a standard forced choice experiment. In fact, most studies of metacognition use two-alternative forced-choice tasks, while decisions in real life are often based on unreliable evidence and are influenced by a number of directly related as well as distant factors (Rahnev et al., 2022). A recent opinion paper has suggested the most important long- and medium-term goals for the field (Rahnev et al., 2022). One of the long-term goals explicitly states the need to translate findings about the mechanisms underlying confidence to more complex settings. In order to accomplish this goal, a first step is to develop tasks with increasing complexity. In our paradigm, we used four response options and showed that the task was feasible for adults across the lifespan and could be used to replicate established findings of behavioural and neural correlates of metacognition. Still, more complex tasks that require, for example, a choice between more than four options, continuous decisions (such as reproducing the orientation of a grating), or multilevel decisions that unfold over time, are conceivable. However, the difficulty of creating sophisticated computational models increases with the number of response alternatives (Rahnev, 2020). Li and Ma (2020) made an effort to gain new insights about the computation of confidence in more complex settings and modelled confidence in a three-alternative categorisation task. They found that confidence did not reflect the subjective likelihood of a correct decision, but rather the chance that the chosen option was more likely than the second best option, thus, the subjective feeling of having chosen the best possible option, irrespective of its absolute probability. The successful development of a computational model explaining confidence judgements in a three-choice task is a first step to apply contemporary models of metacognition to more complex decisions that are closer to real world settings. The generalisability of positive training effects has also been investigated in more complex settings. For instance, metacognition training, a psychological intervention program, successfully reduced symptoms of psychosis and schizophrenia by educating patients about cognitive biases (Eichner & Berna, 2016; Moritz et al., 2010). In the context of ageing, a metacognitive training teaching older adults' skills to solve mathematical problems improved the participants' accuracy of predicting the own performance (Pennequin et al., 2010). These are promising examples of how negative practical implications of an age-related decline in metacognitive performance can be prevented by tailored interventions. More importantly, progress in our understanding of the neurocognitive computation of confidence is needed to explain normal and abnormal impairments related to ageing and might proof valuable to predict the onset of cognitive decline or age-related (neurodegenerative) diseases such as dementia (Wilson et al., 2015).

Coming back to the very beginning, I have portrayed driving on the highway and deciding whether it is safe to pass a car as an example of a complex every-day decision that has serious implications and requires constant monitoring and adjustment of behaviour. The joint findings of this thesis suggest that not only the complex decision-making itself but also its accurate monitoring will become increasingly difficult with older age. Older adults might be highly uncertain about their decisions while driving, which can of course be stressful, but primarily dangerous if wrong decisions are rarely clearly detected as such and will not be avoided in the future. Possibly, external feedback about the driving behaviour (e.g., from a trainer or from the car) might be valuable to improve older adults' decision-making because our findings suggest that they are supposedly able to adjust their behaviour in an adaptive way. From a different point of view, acknowledging the relevance of metacognitive monitoring while driving, a conclusion that could be derived from this thesis could also be to draw on the results of a standard test of metacognitive accuracy (that might be similar to the paradigm used here) when assessing whether a driver's licence should be withdrawn.

4.3. Limitations

While some limitations specific to each study were highlighted throughout the previous chapter, I will mention some general weaknesses and a more theoretical issue here in order to put the findings of this thesis into perspective.

One limitation that should be considered when evaluating the conclusions of this thesis is that all findings were based on data from the same experiment. It remains open whether these findings are general or specific to this experiment (i.e., to the employed paradigm, but also to the setting, the experimenter, etc.). Given that some established findings from the relevant bodies of literature were replicated (e.g., decline in task performance and metacognitive performance and general slowing with older age, negative correlation between confidence and RT, larger $N_{e/c}$ and $P_{e/c}$ amplitudes for errors compared to correct responses) a generalizability of the findings seems likely, but should be tested in replication studies.

Another aspect that could limit the generalizability of our findings is the studied sample. The older adults who participated in this study might not be representative of the general older population because they might have performed above average. This follows from self-selection bias because participants were obviously interested in research and incurred some efforts to register for and attend the experiment. This high-performing sample of older adults might have especially high cognitive reserve, which suggests that the age-related decline in metacognitive performance observed in our study could be less detrimental than in the general population. However, the majority of the younger participants were PhD students and thus also had an above-average education. Despite this potential confound we observed an age-related decline in metacognition, but null findings regarding the effect of age on, for example, the adjustment of behaviour or the modulation of ERP components might be related to this bias.

A general challenge within the study of ageing refers to the interindividual variability in ageing trajectories. Patterns of ageing differ substantially between individuals and evidence shows that certain lifestyle factors (e.g., stress, negative experiences) may in fact have a larger effect on the development of cognitive functions over time than the chronological age (Reuter-Lorenz & Park, 2014). Matching participants across decades regarding a number of factors that could be related to ageing is difficult, but we assessed processing speed (d2-test, Brickenkamp, 2002) and screened for signs of depression (Beck's Depression Inventory, Hautzinger, 1991) and cognitive impairment (Mini-Mental-State Examination, Folstein et al., 1975), ensuring to detect and exclude the extreme cases.

A crucial avenue for future research is to provide a computational account of age-related changes in metacognition. This would add to the findings of this thesis and would be crucial in order to derive a clear picture of which age-related processes are driving the decline in metacognitive performance that we observed, that is, to break down the monitoring process into latent parameters. Computational modelling using variants of the drift diffusion model has provided a detailed account of decision-making of older adults. For example, it was shown that older adults exhibit longer non-decision times (e.g., sensory encoding or motor execution) that lead to the prominent slowing. This is not the only factor though, as older adults also show increased decision bounds, thus applying a more cautious decision policy, which also results in slower responses (Forstmann et al., 2011; Ratcliff et al., 2011; Ratcliff & McKoon, 2015). Applied to second-order decisions, it could be explored, for instance, whether increased uncertainty in older adults can rather be accounted for by modelling age-related adjustments to boundary separation, or by changes in drift rate, if the quality of evidence entering the process of computing confidence is in fact noisier. In case of the former, the question whether this was due to capacity limitations or an intentional strategy, however, would still remain open.

A final and rather theoretical issue pertains to the latent construct behind confidence ratings, that is, what is it that we are actually measuring? This is a recurring question in the field and was recently raised in a recent paper by Caziot and Mamassian (2021). The authors question the prevalent definition of confidence reflecting the subjective probability that the initial decision was correct (Pouget et al., 2016). Instead, they proposed, and confirmed in two sophisticated psychophysical experiments, that it might rather reflect a measure of self-consistency. This means that people generally aim to be self-consistent with themselves and that their metacognitive evaluation of a decision might be related to the evaluation of previous decisions when the same sensory evidence was presented. While this proposal should be considered when interpreting findings, it is generally worth to acknowledge that repetitive confidence judgements in an experiment could be more coherent than in everyday life (Caziot & Mamassian, 2021).

4.4. Summary and conclusion

Metacognition describes the monitoring and evaluation of our thoughts and actions and serves to optimise behaviour by comparing it to the required outcomes. As people age, many cognitive functions decline and more errors occur, which increases the need for effective metacognition.

The goal of this thesis was to examine metacognitive performance in a perceptual conflict task across the lifespan and to identify factors that are related to a potential decline in metacognition by investigating the behavioural and neural correlates of decision confidence. We found that the accuracy of confidence ratings declined significantly across the adult lifespan, while behaviour was nevertheless similarly adjusted according to an internal evaluation of the decision across age. Moreover, we found marked differences in the relationship between confidence ratings and response parameters between younger and older adults. At the neural level, both the $N_{e/c}$ and the $P_{e/c}$ showed larger amplitudes in errors rated with lower confidence. With older age, the sensitivity of the N_e to confidence was reduced, while the P_e indexed the subjective feeling of confidence similar to younger adults.

Overall, these findings reveal multiple factors that are related to an increased uncertainty of older adults about the accuracy of a decision, which might play a role in the computation of confidence. As a result, the accumulation of error evidence seems to be disrupted or noisy in older adults, leading to fewer instances where they are certain that they made an error.

In summary, this thesis advances the understanding of the mechanisms underlying age-related changes in metacognitive performance, even though a unified account describing the effect of the ageing process on the entire process of metacognitive monitoring and evaluation will take us a long way. If future research can provide greater insight into how precisely ageing impacts on metacognitive mechanisms, this would constitute the necessary basis for the development of tailored metacognitive trainings and the identification of early markers for cognitive decline and neurodegenerative diseases. In order to identify which sources of evidence directly affect confidence ratings and how they are weighted, future work should combine the manipulation of several effectors and assess their respective effects on decision confidence across age.

References

Ais, J., Zylberberg, A., Barttfeld, P., & Sigman, M. (2016). Individual consistency in the accuracy and distribution of confidence judgments. *Cognition*, *146*, 377–386. https://doi.org/10.1016/J.COGNITION.2015.10.006

Allen, M., Frank, D., Schwarzkopf, S., Fardo, F., Winston, J. S., Hauser, T. U., & Rees, G. (2016). Unexpected arousal modulates the influence of sensory noise on confidence. *Elife*, *5*(e18103). https://doi.org/10.7554/eLife.18103.001

Anzulewicz, A., Hobot, J., Siedlecka, M., & Wierzchoń, M. (2019). Bringing action into the picture. How action influences visual awareness. *Attention, Perception, and Psychophysics*, *81*(7), 2171–2176. https://doi.org/10.3758/s13414-019-01781-w

Armbrecht, A.-S., Gibbons, H., & Stahl, J. (2012). Monitoring force errors: medial-frontal negativity in a unimanual force-production task. *Psychophysiology*, *49*(1), 56–72. https://doi.org/10.1111/j.1469-8986.2011.01282.x

Armbrecht, A.-S., Gibbons, H., & Stahl, J. (2013). Effects of response force parameters on medial-frontal negativity. *PLoS One*, *8*(1), e54681. https://doi.org/10.1371/journal.pone.0054681

Bäckman, L., Jones, S., Berger, A. K., Laukka, E. J., & Small, B. J. (2005). Cognitive impairment in preclinical Alzheimer’s disease: A meta-analysis. *Neuropsychology*, *19*(4), 520–531. https://doi.org/10.1037/0894-4105.19.4.520

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Learning*, *68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bode, S., & Stahl, J. (2014). Predicting errors from patterns of event-related potentials preceding an overt response. *Biological Psychology*, *103*, 357–369. https://doi.org/10.1016/j.biopsycho.2014.10.002

Boldt, A., de Gardelle, V., & Yeung, N. (2017). The impact of evidence reliability on sensitivity and bias in decision confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(8), 1520–1531.

https://doi.org/10.1037/XHP0000404

Boldt, A., & Yeung, N. (2015). Shared neural markers of decision confidence and error detection. *The Journal of Neuroscience*, *35*(8), 3478–3484. https://doi.org/10.1523/JNEUROSCI.0797-14.2015

Bonder, T., & Gopher, D. (2019). The Effect of Confidence Rating on a Primary Visual Task. *Frontiers in Psychology*, *10*, 2674. https://doi.org/10.3389/FPSYG.2019.02674/BIBTEX

Bonini, F., Burle, B., Lieǵeois-Chauvel, C., Reǵis, J., Chauvel, P., & Vidal, F. (2014). Action monitoring and medial frontal cortex: Leading role of supplementary motor area. *Science*, *343*(6173), 888–891. https://doi.org/10.1126/SCIENCE.1247412/SUPPL_FILE/BONINI.SM.PDF

Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, *402*(6758), 179–181. https://doi.org/10.1038/46035

Brickenkamp, R. (2002). *The d2 Test of attention. (1st ed.)* (9th ed.). Verlag für Psychologie Hogrefe.

Brooks, M. E., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Machler, M., & Bolker, B. M. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, *9*(2), 378–399. https://doi.org/10.3929/ethz-b-000240890

Bunce, D., Macdonald, S. W. S., & Hultsch, D. F. (2004). Inconsistency in serial choice decision and motor reaction times dissociate in younger and older adults q. *Brain and Cognition*, *56*, 320–327. https://doi.org/10.1016/j.bandc.2004.08.006

Cabeza, R. (2001). Cognitive neuroscience of aging: Contributions of functional neuroimaging. *Scandinavian Journal of Psychology*, *42*(3), 277–286. https://doi.org/10.1111/1467-9450.00237

Carpenter, J., Sherman, M. T., Kievit, R. A., Seth, A. K., Lau, H., & Fleming, S. M. (2019). Domain-general enhancements of metacognitive ability through adaptive training.

Journal of Experimental Psychology: General, *148*(1), 51–64. https://doi.org/https://doi.org/10.1037/xge0000505

Castel, A. D., Middlebrooks, C. D., & McGillivray, S. (2016). Monitoring Memory in Old Age: Impaired, Spared, and Aware. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford Handbook of Metamemory* (pp. 519–534). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.3

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421. https://doi.org/10.1016/j.tics.2014.04.012

Caziot, B., & Mamassian, P. (2021). Perceptual confidence judgments reflect self-consistency. *Journal of Vision*, *21*(12), 8–8. https://doi.org/10.1167/JOV.21.12.8

Chan, Y. M., Pianta, M. J., Bode, S., & McKendrick, A. M. (2017). Spatio-temporal patterns of event-related potentials related to audiovisual synchrony judgments in older adults. *Neurobiology of Aging*, *55*, 38–48. https://doi.org/10.1016/j.neurobiolaging.2017.03.011

Charles, L., Van Opstal, F., Marti, S., & Dehaene, S. (2013). Distinct brain mechanisms for conscious versus subliminal error detection. *NeuroImage*, *73*, 80–94. https://doi.org/10.1016/j.neuroimage.2013.01.054

Charles, L., & Yeung, N. (2019). Dynamic sources of evidence supporting confidence judgments and error detection. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(1), 39–52. https://doi.org/10.1037/xhp0000583

Chua, E., Schacter, D. L., & Sperling, R. (2009). Neural basis for recognition confidence. *Psychology and Aging*, *24*(1), 139–153. https://doi.org/10.1037/a0014029.Neural

Clayson, P. E., Baldwin, S. A., & Larson, M. J. (2013). How does noise affect amplitude and latency measurement of event-related potentials (ERPs)? A methodological critique and simulation study. *Psychophysiology*, *50*(2), 174–186. https://doi.org/10.1111/psyp.12001

Cohen, M. X., & van Gaal, S. (2014). Subthreshold muscle twitches dissociate oscillatory neural signatures of conflicts from errors. *NeuroImage*, *86*, 503–513. https://doi.org/10.1016/j.neuroimage.2013.10.033

Cohen, M. X., van Gaal, S., Ridderinkhof, K. R., & Lamme, V. A. F. (2009). Unconscious errors enhance prefrontal-occipital oscillatory synchrony. *Frontiers in Human Neuroscience*, *3*(NOV), 54. https://doi.org/10.3389/NEURO.09.054.2009/BIBTEX

Coles, M. G. H., Scheffers, M. K., & Holroyd, C. B. (2001). Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing. *Biological Psychology*, *56*(3), 173–189. https://doi.org/10.1016/S0301-0511(01)00076-X

Cosentino, S., Metcalfe, J., Butterfield, B., & Stern, Y. (2007). Objective Metamemory Testing Captures Awareness of Deficit in Alzheimer's Disease. *Cortex*, *43*(7), 1004–1019. https://doi.org/10.1016/S0010-9452(08)70697-X

Crone, N., Miglioretti, D. L., Gordon, B., & Lesser, R. P. (1998). Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. I. Alpha and beta event-related desynchronization. *Brain: A Journal of Neurology*, *121*(12), 2271–2299. https://doi.org/10.1093/brain/121.12.2271

Curran, T., Hills, A., Patterson, M. B., & Strauss, M. E. (2001). Effects of aging on visuospatial attention: an ERP study. *Neuropsychologia*, *39*(3), 288–301. https://doi.org/10.1016/S0028-3932(00)00112-3

Danielmeier, C., & Ullsperger, M. (2011). Post-error adjustments. *Frontiers in Psychology*, *2*, 1–10.

Desender, K., Boldt, A., Verguts, T., & Donner, T. H. (2019). Confidence predicts speed-accuracy tradeoff for subsequent decisions. *ELife*, *8*. https://doi.org/10.1101/466730

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective Confidence Predicts Information Seeking in Decision Making. *Psychological Science*, *29*(5), 761–778. https://doi.org/10.1177/0956797617744771

Desender, K., Donner, T. H., & Verguts, T. (2021). Dynamic expressions of confidence within an evidence accumulation framework. *Cognition*, *207*, 104522.

Desender, K., Murphy, P. R., Boldt, A., Verguts, T., & Yeung, N. (2019). A Postdecisional Neural Marker of Confidence Predicts Information-Seeking in Decision-Making. *Journal of Neuroscience*, *39*(17), 3309–3319. https://doi.org/10.1101/433276

Desender, K., Ridderinkhof, K. R., & Murphy, P. R. (2021). Understanding neural signals of post-decisional performance monitoring: An integrative review. *ELife*, *10*. https://doi.org/10.7554/eLife.67556

Di Gregorio, F., Maier, M. E., & Steinhauser, M. (2018). Errors can elicit an error positivity in the absence of an error negativity: Evidence for independent systems of human error monitoring. *NeuroImage*, *172*, 427–436. https://doi.org/10.1016/j.neuroimage.2018.01.081

Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Aging, metamemory, and high-confidence errors: A misrecollection account. *Psychology and Aging*, *22*(1), 122–133. https://doi.org/10.1037/0882-7974.22.1.122

Dotan, D., Meyniel, F., & Dehaene, S. (2018). On-line confidence monitoring during decision making. *Cognition*, *171*(171), 112–121. https://doi.org/10.1016/j.cognition.2017.11.001

Dully, J., McGovern, D. P., & O'Connell, R. G. (2018). The impact of natural aging on computational and neural indices of perceptual decision making: A review. *Behavioural Brain Research*, *355*, 48–55. https://doi.org/10.1016/j.bbr.2018.02.001

Dutilh, G., Forstmann, B. U., Vandekerckhove, J., & Wagenmakers, E. J. (2013). A diffusion model account of age differences in posterror slowing. *Psychology and Aging*, *28*(1), 64–76. https://doi.org/10.1037/A0029875

Eichner, C., & Berna, F. (2016). Acceptance and Efficacy of Metacognitive Training (MCT) on Positive Symptoms and Delusions in Patients With Schizophrenia: A Meta-analysis Taking Into Account Important Moderators. *Schizophrenia Bulletin*, *42*(4), 952–962. https://doi.org/10.1093/SCHBUL/SBV225

Eimer, M., & Coles, M. G. H. (2003). The Lateralized Readiness Potential. In *The Bereitschaftspotential* (pp. 229–248). Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-0189-3_14

Endrass, T., Klawohn, J., Preuss, J., & Kathmann, N. (2012). Temporospatial dissociation of Pe subcomponents for perceived and unperceived errors. *Frontiers in Human Neuroscience, 6*, 1–10. https://doi.org/10.3389/fnhum.2012.00178

Endrass, T., Schreiber, M., & Kathmann, N. (2012). Speeding up older adults: Age-effects on error processing in speed and accuracy conditions. *Biological Psychology, 89*(2), 426–432. https://doi.org/10.1016/j.biopsycho.2011.12.005

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149.

Faivre, N., Filevich, E., Solovey, G., Kühn, S., & Blanke, O. (2018). Behavioral, modeling, and electrophysiological evidence for supramodality in human metacognition. *Journal of Neuroscience*, *38*(2), 263–277. https://doi.org/10.1523/JNEUROSCI.0322-17.2017

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology*, *78*(6), 447–455. https://doi.org/10.1016/0013-4694(91)90062-9

Falkenstein, M., Hoormann, J., & Hohnsbein, J. (2001). Changes of error-related ERPs with age. *Experimental Brain Research*, *138*(2), 258–262. https://doi.org/10.1007/s002210100712

Falkenstein, M., Yordanova, J., & Kolev, V. (2006). Effects of aging on slowing of motor-response generation. *International Journal of Psychophysiology*, *59*(1), 22–29. https://doi.org/10.1016/J.IJPSYCHO.2005.08.004

Feuerriegel, D., Murphy, M., Konski, A., Mepani, V., Sun, J., Hester, R., & Bode, S. (2022). Electrophysiological correlates of confidence differ across correct and erroneous perceptual decisions. *NeuroImage*, *259*(November 2021), 119447. https://doi.org/10.1016/j.neuroimage.2022.119447

Ficarella, S. C., Rochet, N., & Burle, B. (2019). Becoming aware of subliminal responses: An EEG/EMG study on partial error detection and correction in humans. *Cortex*, *120*, 443–456. https://doi.org/10.1016/j.cortex.2019.07.007

Filevich, E., Koß, C., & Faivre, N. (2020). Response-related signals increase confidence but not metacognitive performance. *ENeuro*, *7*(3), 1–14. https://doi.org/10.1523/ENEURO.0326-19.2020

Filippi, R., Ceccolini, A., Periche-tomas, E., & Bright, P. (2020). Developmental trajectories of metacognitive processing and executive function from childhood to older age. *Quarterly Journal of Experimental Psychology*, *73*(11), 1757–1773. https://doi.org/10.1177/1747021820931096

Fitzgerald, L. M., Arvaneh, M., & Dockree, P. M. (2017). Domain-specific and domain-general processes underlying metacognitive judgments. *Consciousness and Cognition*, *49*, 264–277. https://doi.org/10.1016/j.concog.2017.01.011

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906. https://psycnet.apa.org/record/1980-09388-001

Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review*, *124*(1), 91–114. https://doi.org/10.1037/REV0000045

Fleming, S. M., & Dolan, R. J. (2010). Effects of loss aversion on post-decision wagering: Implications for measures of awareness. *Consciousness and Cognition*, *19*(1), 352–363. https://doi.org/10.1016/J.CONCOG.2009.11.002

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1338–1349. https://doi.org/10.1098/rstb.2011.0417

Fleming, S. M., Dolan, R. J., & Frith, C. D. (2012). Metacognition: Computation, biology and function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1280–1286. https://doi.org/10.1098/rstb.2012.0021

Fleming, S. M., & Frith, C. D. (2014). The Cognitive Neuroscience of Metacognition. In *The Cognitive Neuroscience of Metacognition*. Springer. https://doi.org/10.1007/978-3-642-45190-4

Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012). Prefrontal contributions to metacognition in perceptual decision making. *Journal of Neuroscience, 32*(18), 6117–6125. https://doi.org/10.1523/JNEUROSCI.6489-11.2012

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience, 8*, 1–9. https://doi.org/10.3389/fnhum.2014.00443

Fleming, S. M., Maniscalco, B., Ko, Y., Amendi, N., Ro, T., & Lau, H. (2015). Action-Specific Disruption of Perceptual Confidence. *Psychological Science*, *26*(1), 89–98. https://doi.org/10.1177/0956797614557697

Fleming, S. M., Massoni, S., Gajdos, T., & Vergnaud, J.-C. (2016). Metacognition about the past and future: quantifying common and distinct influences on prospective and retrospective judgments of self-performance. *Neuroscience of Consciousness, 2016*(1), niw018. https://doi.org/10.1093/NC/NIW018

Fleming, S. M., Van Der Putten, E. J., & Daw, N. D. (2018). Neural mediators of changes of mind about perceptual decisions. *Nature Neuroscience*, *21*(4), 617–624. https://doi.org/10.1038/s41593-018-0104-6

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science*, *329*(5998), 1541–1543. https://doi.org/10.1126/SCIENCE.1191883

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Forstmann, B. U., Tittgemeyer, M., Wagenmakers, E. J., Derrfuss, J., Imperati, D., & Brown, S. (2011). The Speed-Accuracy Tradeoff in the Elderly Brain: A Structural Model-Based Approach. *Journal of Neuroscience*, *31*(47), 17242–17249. https://doi.org/10.1523/JNEUROSCI.0309-11.2011

Franz, E. A., & Miller, J. (2002). Effects of response readiness on reaction time and force output in people with Parkinson's disease. *Brain*, *125*(8), 1733–1750. https://doi.org/10.1093/brain/awf192

Gajdos, T., Fleming, S. M., Saez Garcia, M., Weindel, G., & Davranche, K. (2019). Revealing subthreshold motor contributions to perceptual confidence. *Neuroscience of Consciousness*, *5*(1), 1–8. https://doi.org/10.1093/nc/niz001

Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. *Psychonomic Bulletin and Review*, *10*(4), 843–876. https://doi.org/10.3758/BF03196546

Garavan, H., Hester, R., Murphy, K., Fassbender, C., & Kelly, C. (2006). Individual differences in the functional neuroanatomy of inhibitory control. *Brain Research*, *1105*(1), 130–142. https://doi.org/10.1016/J.BRAINRES.2006.03.029

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A Neural System for Error Detection and Compensation. *Psychological Science*, *4*(6), 385–390. https://doi.org/10.1111/j.1467-9280.1993.tb00586.x

Gherman, S., & Philiastides, M. G. (2015). Neural representations of confidence emerge from the process of decision formation during perceptual choices. *NeuroImage*, *106*, 134–143. https://doi.org/https://doi.org/10.1016/j.neuroimage.2014.11.036

Gold, J., neuroscience, M. S.-A. review of, & 2007, undefined. (2007). The neural basis of decision making. *Sunyopt.Edu*. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Grady, C. (2012). Trends in Neurocognitive Aging. *Nature Reviews Neuroscience*, *13*(7), 491–505. https://doi.org/10.1038/nrn3256

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Wiley.

Groom, M. J., & Cragg, L. (2015). Differential modulation of the N2 and P3 event-related potentials by response conflict and inhibition. *Brain and Cognition*, *97*, 1–9. https://doi.org/10.1016/j.bandc.2015.04.004

Hagura, N., Haggard, P., & rn Diedrichsen, J. (2017). Perceptual decisions are biased by the cost to act. *ELife*, *6*, 18422. https://doi.org/10.7554/eLife.18422.001

Hajcak, G., McDonald, N., & Simons, R. F. (2003). To err is autonomic: Error-related brain potentials, ANS activity, and post-error compensatory behavior. *Psychophysiology*, *40*(6), 895–903. https://doi.org/10.1111/1469-8986.00107

Hajcak, G., & Simons, R. F. (2002). Error-related brain activity in obsessive–compulsive undergraduates. *Psychiatry Research*, *110*(1), 63–72. https://doi.org/10.1016/S0165-1781(02)00034-3

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research*, *18*(10), 1279–1296. https://doi.org/10.1016/0042-6989(78)90218-3

Hämmerer, D., Müller, V., & Li, S. C. (2014). Performance monitoring across the lifespan: Still maturing post-conflict regulation in children and declining task-set monitoring in older adults. *Neuroscience and Biobehavioral Reviews*, *46*, 105–123. https://doi.org/10.1016/j.neubiorev.2014.06.008

Hansson, P., Rönnlund, M., Juslin, P., & Nilsson, L. G. (2008). Adult Age Differences in the Realism of Confidence Judgments: Overconfidence, Format Dependence, and Cognitive Predictors. *Psychology and Aging*, *23*(3), 531–544. https://doi.org/10.1037/a0012782

Harada, C. N., Natelson Love, M. C., & Triebel, K. L. (2013). Normal Cognitive Aging. *Clinics in Geriatric Medicine*, *29*(4), 737–752. https://doi.org/10.1016/J.CGER.2013.07.002

Harty, S., Murphy, P. R., Robertson, I. H., & O’Connell, R. G. (2017). Parsing the neural signatures of reduced error detection in older age. *NeuroImage*, *161*, 43–55. https://doi.org/10.1016/j.neuroimage.2017.08.032

Harty, S., O’Connell, R. G., Hester, R., & Robertson, I. H. (2013). Older adults have diminished awareness of errors in the laboratory and daily life. *Psychology and Aging*, *28*(4), 1032–1041. https://doi.org/10.1037/a0033567

Hautzinger, M. (1991). The Beck Depression Inventory in clinical practice. *Der Nervenarzt*, *62*(11), 689–696.

Hecht, H., Vogt, S., & Prinz, W. (2001). Motor learning enhances perceptual judgment: a case for action-perception transfer. *Psychological Research*, *65*(1), 3–14.

https://doi.org/10.1007/s004260000043

Hertzog, C., & Curley, T. (2018). Metamemory and Cognitive Aging. In *Oxford Research Encyclopedia of Psychology*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190236557.013.377

Hertzog, C., & Dunlosky, J. (2011). Metacognition in Later Adulthood: Spared Monitoring Can Benefit Older Adults' Self-regulation. *Current Directions in Psychological Science*, *20*(3), 167–173. https://doi.org/10.1177/0963721411409026

Hertzog, C., & Pearman, A. (2013). Memory Complaints in Adulthood and Old Age. In *The SAGE handbook of applied memory* (pp. 423–444).

Heuninckx, S., Wenderoth, N., & Swinnen, S. P. (2008). Systems Neuroplasticity in the Aging Brain: Recruiting Additional Neural Resources for Successful Motor Performance in Elderly Persons. *Journal of Neuroscience*, *28*(1), 91–99. https://doi.org/10.1523/JNEUROSCI.3300-07.2008

Heyes, C., Bang, D., Shea, N., Frith, C. D., & Fleming, S. M. (2020). Knowing Ourselves Together: The Cultural Origins of Metacognition. *Trends in Cognitive Sciences*, *24*(5), 349–362. https://doi.org/10.1016/j.tics.2020.02.007

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679–709. https://doi.org/10.1037/0033-295X.109.4.679

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: with applications in R*. Springer.

Jaśkowski, P., Van Der Lubbe, R. H. J., Wauschkuhn, B., Wascher, E., & Verleger, R. R. (2000). The influence of time pressure and cue validity on response force in an S1-S2 paradigm. *Acta Psychologica*, *105*(1), 89–105. https://doi.org/10.1016/S0001-6918(00)00046-9

Kantowitz, B. H. (1973). Response force as an indicant of conflict in double stimulation. *Journal of Experimental Psychology*, *100*(2), 302–309. https://doi.org/10.1037/h0035780

Katya, S., & Fleming, S. M. (2022). *Bayesian meta-analysis reveals probable increase in metacognitive efficiency through feedback Metacognitive efficiency*. [Poster] ASSC 25, Amsterdam. https://theassc.org/assc-25/

Kayser, J., & Tenke, C. E. (2015). On the benefits of using surface Laplacian (Current Source Density) methodology in electrophysiology. *Journal of Psychophysiology*, *97*(3), 171–173. https://doi.org/10.1016/j.ijpsycho.2015.06.001

Kepecs, A., & Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1322–1337. https://doi.org/10.1098/RSTB.2012.0037

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature 2008 455:7210*, *455*(7210), 227–231. https://doi.org/10.1038/nature07200

Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron*, *84*(6), 1329–1342. https://doi.org/10.1016/j.neuron.2014.12.015

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, *324*(5928), 759–764. https://doi.org/10.1126/SCIENCE.1169405

Klawohn, J., Santopetro, N. J., Meyer, A., & Hajcak, G. (2020). Reduced P300 in depression: Evidence from a flanker task and impact on ERN, CRN, and Pe. *Psychophysiology*, *57*(4), 1–11. https://doi.org/10.1111/psyp.13520

Ko, Y. H., Feuerriegel, D., Turner, W., Overhoff, H., Niessen, E., Stahl, J., Hester, R., Fink, G. R., Weiss, P. H., & Bode, S. (2022). Divergent effects of absolute evidence magnitude on decision accuracy and confidence in perceptual judgements. *Cognition*, *225*, 105125. https://doi.org/10.1016/j.cognition.2022.105125

Ko, Y. T., Alsford, T., & Miller, J. (2012). Inhibitory effects on response force in the stop-signal paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 465–477. https://doi.org/10.1037/a0027034

Kolev, V., Falkenstein, M., & Yordanova, J. (2005). Aging and error processing: Time-frequency analysis of error-related potentials. *Journal of Psychophysiology*, *19*(4), 289–297. https://doi.org/10.1027/0269-8803.19.4.289

Kornell, N., Son, L. K., & Terrace, H. S. (2007). Transfer of metacognitive skills and hint seeking in monkeys: Research article. *Psychological Science*, *18*(1), 64–71. https://doi.org/10.1111/j.1467-9280.2007.01850.x

Korsch, M., Frühholz, S., & Herrmann, M. (2016). Conflict-specific aging effects mainly manifest in early information processing stages-an ERP study with different conflict types. *Frontiers in Aging Neuroscience*, *8*, 1–12. https://doi.org/10.3389/fnagi.2016.00053

Kruskal, W. H., & Goodman, L. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, *49*(268), 732–764.

Larson, M. J., Clayson, P. E., Keith, C. M., Hunt, I. J., Hedges, D. W., Nielsen, B. L., & Call, V. R. A. (2016). Cognitive control adjustments in healthy older and younger adults: Conflict adaptation, the error-related negativity (ERN), and evidence of generalized decline with age. *Biological Psychology*, *115*, 50–63. https://doi.org/10.1016/j.biopsycho.2016.01.008

Latash, M. (2021). Efference copy in kinesthetic perception: a copy of what is it? *Journal of Neurophysiology*, *125*, 1079–1094. https://doi.org/10.1152/jn.00545.2020

Lau, H. C., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(49), 18763–18768. https://doi.org/10.1073/PNAS.0607716103

Lehmann, M., Hagen, J., & Ettinger, U. (2022). Unity and diversity of metacognition. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001197

Lezak, M. D., Howieson, D. B., Loring, D. W., & Fischer, J. S. (Eds.). (2004). *Neuropsychological assessment*. Oxford University Press.

Li, H. H., & Ma, W. J. (2020). Confidence reports in decision-making with multiple alternatives violate the Bayesian confidence hypothesis. *Nature Communications*, *11*(1). https://doi.org/10.1038/S41467-020-15581-6

Long, J. A. (2019). *interactions: Comprehensive, User-Friendly Toolkit for Probing Interactions* (1.1.0).

Lucci, G., Berchicci, M., Spinelli, D., Taddei, F., & Di Russo, F. (2013). The Effects of Aging on Conflict Detection. *PLoS ONE*, *8*(2). https://doi.org/10.1371/journal.pone.0056566

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique* (S. J. Luck (Ed.)). MIT press.

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software, 6*(60), 3139. https://doi.org/10.21105/joss.03139

Maier, M. E., & Steinhauser, M. (2017). Working memory load impairs the evaluation of behavioral errors in the medial frontal cortex. *Psychophysiology*, *54*(10), 1472–1482. https://doi.org/10.1111/psyp.12899

Maier, M. E., Steinhauser, M., & Hübner, R. (2008). Is the Error-related Negativity Amplitude Related to Error Detectability? Evidence from Effects of Different Error Types. *Journal of Cognitive Neuroscience*, *20*(12), 2263–2273.

Maier, M. E., Steinhauser, M., & Hübner, R. (2010). Effects of response-set size on error-related brain activity. *Experimental Brain Research*, *202*(3), 571–581. https://doi.org/10.1007/s00221-010-2160-3

Maillet, D., Reviews, M. R.-A. research, & 2013, U. (2013). Association between prefrontal activity and volume change in prefrontal and medial temporal lobes in aging and dementia: a review. *Ageing Research Reviews*, *12*, 479–489. https://www.sciencedirect.com/science/article/pii/S156816371200147X

Mamassian, P. (2020). Confidence forced-choice and other metaperceptual tasks. *Perception*, *49*(6), 616–635. https://doi.org/10.1177/0301006620928010

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*(1), 422–430. https://doi.org/10.1016/j.concog.2011.09.021

Masina, F., Di Rosa, E., & Mapelli, D. (2018). Intra-individual variability of error awareness and post-error slowing in three different age-groups. *Frontiers in Psychology*, *9*, 1–10. https://doi.org/10.3389/fpsyg.2018.00902

Masliah, E., Mallory, M., Hansen, L., DeTeresa, R., & Terry, R. D. (1993). Quantitative synaptic alterations in the human neocortex during normal aging. *Neurology*, *43*, 192–192. https://doi.org/10.1212/WNL.43.1_PART_1.192

Mathewson, K. J., Dywan, J., & Segalowitz, S. J. (2005). Brain bases of error-related ERPs as influenced by age and task. *Biological Psychology*, *70*(2), 88–104. https://doi.org/10.1016/j.biopsycho.2004.12.005

Mattes, A., Porth, E., & Stahl, J. (2022). Linking neurophysiological processes of action monitoring to post-response speed-accuracy adjustments in a neuro-cognitive diffusion model. *NeuroImage*, *247*. https://doi.org/10.1016/j.neuroimage.2021.118798

Mattes, S., & Ulrich, R. (1997). Response force is sensitive to the temporal uncertainty of response stimuli. *Perception and Psychophysics*, *59*(7), 1089–1097. https://doi.org/10.3758/BF03205523

Mattes, S., Ulrich, R., & Miller, J. (2002). Response force in RT tasks: Isolating effects of stimulus probability and response probability. *Visual Cognition*, *9*(4–5), 477–501. https://doi.org/10.1080/13506280143000548

McGillivray, S., & Castel, A. D. (2011). Betting on Memory Leads to Metacognitive Improvement by Younger and Older Adults. *Psychology and Aging*, *26*(1), 137–142. https://doi.org/10.1037/A0022681

McWilliams, A., Bibby, H., Steinbeis, N., David, A. S., & Fleming, S. M. (2022). Age-related decreases in global metacognition are independent of local metacognition and task performance. *PsyArXiv*, 1–23.

Mecacci, L., & Righi, S. (2006). Cognitive failures, metacognitive beliefs and aging.

Personality and Individual Differences, *40*(7), 1453–1459. https://doi.org/10.1016/j.paid.2005.11.022

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review 2008 15:1*, *15*(1), 174–179. https://doi.org/10.3758/PBR.15.1.174

Middlebrooks, P. G., & Sommer, M. A. (2011). Metacognition in monkeys during an oculomotor task. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(2), 325–337. https://doi.org/10.1037/A0021611

Moran, R., Teodorescu, A. R., & Usher, M. (2015). Post choice information integration as a causal determinant of confidence: Novel data and a computational account. *Cognitive Psychology*, *78*, 99–147. https://doi.org/10.1016/j.cogpsych.2015.01.002

Morcom, A. M., & Henson, R. N. A. (2018). Increased Prefrontal Activity with Aging Reflects Nonspecific Neural Responses Rather than Compensation. *Journal of Neuroscience*, *38*(33), 7303–7313. https://doi.org/10.1523/JNEUROSCI.1701-17.2018

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes Factors for Common Designs.* (R package version 0.9.12-4.2). https://cran.r-project.org/package=BayesFactor

Moritz, S., Vitzthum, F., Randjbar, S., Veckenstedt, R., & Woodward, T. S. (2010). Detecting and defusing cognitive traps: Metacognitive intervention in schizophrenia. *Current Opinion in Psychiatry*, *23*(6), 561–569. https://doi.org/10.1097/YCO.0B013E32833D16A8

Murman, D. L. (2015). The Impact of Age on Cognition. *Seminars in Hearing*, *36*(3), 111. https://doi.org/10.1055/S-0035-1555115

Murphy, P. R., Robertson, I. H., Harty, S. S., O’Connell, R. G., & O’Connell, R. G. (2015). Neural evidence accumulation persists after choice to inform metacognitive judgments. *ELife*, *4*, 1–23. https://doi.org/10.7554/eLife.11946

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*(1), 109–133.

https://doi.org/https://doi.org/10.1037/0033-2909.95.1.109

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 125–144). Academic Press.

Nguyen, T. Van, Hsu, C. Y., Jaiswal, S., Muggleton, N. G., Liang, W. K., & Juan, C. H. (2021). To Go or Not to Go: Degrees of Dynamic Inhibitory Control Revealed by the Function of Grip Force and Early Electrophysiological Indices. *Frontiers in Human Neuroscience*, *15*. https://doi.org/10.3389/fnhum.2021.614978

Niessen, E., Fink, G. R., Hoffmann, H. E. M. M., Weiss, P. H., & Stahl, J. (2017). Error detection across the adult lifespan: Electrophysiological evidence for age-related deficits. *NeuroImage*, *152*, 517–529. https://doi.org/10.1016/j.neuroimage.2017.03.015

Nieuwenhuis, S., Richard Ridderinkhof, K., Blom, J., Band, G. P. H., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology*, *38*(5), 752–760. https://doi.org/10.1017/S0048577201001111

Notebaert, W., Houtman, F., Opstal, F. Van, Gevers, W., Fias, W., & Verguts, T. (2009). Post-error slowing: An orienting account. *Cognition*, *111*(2), 275–279. https://doi.org/10.1016/J.COGNITION.2009.02.002

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*(12), 1729–1735. https://doi.org/10.1038/nn.3248

O'Shea, J., Sebastian, C., Boorman, E. D., Johansen-Berg, H., & Rushworth, M. F. S. (2007). Functional specificity of human premotor-motor cortical interactions during action selection. *European Journal of Neuroscience*, *26*(7), 2085–2095. https://doi.org/10.1111/j.1460-9568.2007.05795.x

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Overhoff, H., Ko, Y. H., Feuerriegel, D., Fink, G. R., Stahl, J., Weiss, P. H., Bode, S., & Niessen, E. (2021). Neural correlates of metacognition across the adult lifespan. *Neurobiology of Aging*, *108*, 34–46. https://doi.org/10.1016/j.neurobiolaging.2021.08.001

Palmer, E. C., David, A. S., & Fleming, S. M. (2014). Effects of age on metacognitive efficiency. *Consciousness and Cognition*, *28*(1), 151–160. https://doi.org/10.1016/j.concog.2014.06.007

Palser, E. R., Fotopoulou, A., & Kilner, J. M. (2018). Altering movement parameters disrupts metacognitive accuracy. *Consciousness and Cognition*, *57*(November 2017), 33–40. https://doi.org/10.1016/j.concog.2017.11.005

Pansky, A., Goldsmith, M., Koriat, A., & Pearlman-Avnion, S. (2009). Memory accuracy in old age: Cognitive, metacognitive, and neurocognitive determinants. *European Journal of Cognitive Psychology*, *21*(2–3), 303–329. https://doi.org/10.1080/09541440802281183

Parés-Pujolràs, E., Hatchuel, J., & Haggard, P. (2020). A neural turning point - the EEG P3 component tracks unfolding changes of mind. *BioRxiv*, 2020.11.18.388363. https://doi.org/10.1101/2020.11.18.388363

Park, J., Carp, J., Hebrank, A., Park, D. C., & Polk, T. A. (2010). Neural Specificity Predicts Fluid Processing Ability in Older Adults. *The Journal of Neuroscience*, *30*(27), 9253–9259. https://doi.org/10.1523/JNEUROSCI.0853-10.2010

Pennequin, V., Sorel, O., & Mainguy, M. (2010). Metacognition, executive functions and aging: The effect of training in the use of metacognitive skills to solve mathematical word problems. *Journal of Adult Development*, *17*(3), 168–176. https://doi.org/10.1007/s10804-010-9098-3

Pereira, M., Faivre, N., Iturrate, I., Wirthlin, M., Serafini, L., Martin, S., Desvachez, A., Blanke, O., de Ville, D. Van, & del Millán, J. R. (2020). Disentangling the origins of confidence in speeded perceptual judgments through multimodal imaging. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(15), 8382–8390. https://doi.org/10.1073/pnas.1918335117

Peters, M. A. K. (2022). Confidence in Decision-Making. *Neuroscience Oxford Research Encyclopedias*, 1–21. https://oxfordre.com/neuroscience/view/10.1093/acrefore/9780190264086.001.0001/acrefore-9780190264086-e-371

Peters, M. A. K., Thesen, T., Ko, Y. D., Maniscalco, B., Carlson, C., Davidson, M., Doyle, W., Kuzniecky, R., Devinsky, O., Halgren, E., & Lau, H. (2017). Perceptual confidence neglects decision-incongruent evidence in the brain. *Nature Human Behaviour*, *1*(7), 1–21. https://doi.org/10.1038/s41562-017-0139

Pietschmann, M., Endrass, T., & Kathmann, N. (2011). Age-related alterations in performance monitoring during and after learning. *Neurobiology of Aging*, *32*(7), 1320–1330. https://doi.org/10.1016/J.NEUROBIOLAGING.2009.07.016

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, *117*(3), 864–901. https://doi.org/10.1037/A0019737

Polich, J. (1996). Meta-analysis of P300 normative aging studies. *Psychophysiology*, *33*(4), 334–353. https://doi.org/10.1111/J.1469-8986.1996.TB01058.X

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: distinct probabilistic quantities for different goals. *Nature Neuroscience*, *19*(3), 366–374. https://doi.org/10.1038/nn.4240

Quandt, F., Bönstrup, M., Schulz, R., Timmermann, J. E., Zimerman, M., Nolte, G., & Hummel, F. C. (2016). Spectral Variability in the Aged Brain during Fine Motor Control. *Frontiers in Aging Neuroscience*, *8*, 1–9. https://doi.org/10.3389/FNAGI.2016.00305

R Core Team. (2021). *R: A language and environment for statistical computing* (4.0.5). R Foundation for Statistical Computing.

Rabbitt, P. M. A. (1966). Errors and error correction in choice-response tasks. *Experimental*

Psychology, *71*(2), 264–272. https://doi.org/10.1037/h0022853

Rabbitt, P. M. A. (1967). Time to detect errors as a function of factors affecting choice-response time. *Acta Psychologica*, *27*, 131–142. https://www.sciencedirect.com/science/article/pii/0001691867900534

Rabbitt, P. M. A. (1990). Age, IQ and awareness, and recall of errors. *Ergonomics*, *33*(10–11), 1291–1305. https://doi.org/10.1080/00140139008925333

Rahnev, D. (2020). Confidence in the Real World. *Trends in Cognitive Sciences*, *24*(8), 590–591. https://doi.org/10.1016/j.tics.2020.05.005

Rahnev, D., Balsdon, T., Charles, L., Gardelle, V. de, Denison, R., Desender, K., Faivre, N., Filevich, E., Fleming, S. M., Jehee, J., Lau, H., Lee, A. L. F., Locke, S. M., Mamassian, P., Odegaard, B., Peters, M., Reyes, G., Rouault, M., Sackur, J., … Zylberberg, A. (2022). Consensus Goals in the Field of Visual Metacognition. *Perspectives on Psychological Science*. https://doi.org/https://doi.org/10.1177/17456916221075615

Rahnev, D., Desender, K., Lee, A. L. F., Adler, W. T., Aguilar-Lleyda, D., Akdoğan, B., Arbuzova, P., Atlas, L. Y., Balcı, F., Bang, J. W., Bègue, I., Birney, D. P., Brady, T. F., Calder-Travis, J., Chetverikov, A., Clark, T. K., Davranche, K., Denison, R. N., Dildine, T. C., … Zylberberg, A. (2020). The Confidence Database. *Nature Human Behaviour*, *4*(3), 317–325. https://doi.org/10.1038/s41562-019-0813-1

Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/NECO.2008.12-06-420

Ratcliff, R., & McKoon, G. (2015). Aging effects in item and associative recognition memory for pictures and words. *Psychology and Aging*, *30*(3), 669–674. https://doi.org/10.1037/PAG0000030

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, *20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology: General*, *140*(3), 464–487. https://doi.org/10.1037/A0023810

Rausch, M., Zehetleitner, M., Steinhauser, M., & Maier, M. E. (2020). Cognitive modelling reveals distinct electrophysiological markers of decision confidence and error monitoring. *NeuroImage*, *218*, 116963. https://doi.org/10.1016/J.NEUROIMAGE.2020.116963

Reitan, R. M. (1992). *Trail Making Test. Manual for administration and scoring*. Reitan Neurological Laboratory.

Resnick, S. M., Pham, D. L., Kraut, M. A., Zonderman, A. B., & Davatzikos, C. (2003). Longitudinal Magnetic Resonance Imaging Studies of Older Adults: A Shrinking Brain. *Journal of Neuroscience*, *23*(8), 3295–3301. https://doi.org/10.1523/JNEUROSCI.23-08-03295.2003

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, *461*(7261), 263–266. https://doi.org/10.1038/nature08275

Reuter-Lorenz, P. A., Marshuetz, C., Jonides, J., Smith, E. E., Hartley, A., & Koeppe, R. (2001). Neurocognitive ageing of storage and executive processes. *European Journal of Cognitive Psychology*, *13*(1–2), 257–278. https://doi.org/10.1080/09541440125972

Reuter-Lorenz, P. A., & Park, D. C. (2014). How Does it STAC Up? Revisiting the Scaffolding Theory of Aging and Cognition. *Neuropsychology Review*, *24*, 355–370. https://doi.org/10.1007/s11065-014-9270-9

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, *306*(5695), 443–447. https://doi.org/10.1126/science.1100301

Rosi, A., Cavallini, E., Iannello, P., Antonietti, A., & Hertzog, C. (2018). Personal and general beliefs about decision-making in young, middle-age and older adults. *Cognitive Aging Conference*, 271–271. https://publicatt.unicatt.it/handle/10807/121170

Rosjat, N., Wang, B. A., Liu, L., Fink, G. R., & Daun, S. (2020). Stimulus transformation into motor action: Dynamic graph analysis reveals a posterior-to-anterior shift in brain network communication of older subjects. *Human Brain Mapping*, *42*(5), 1547–1563. https://doi.org/10.1002/hbm.25313

Ross, L. A., Dodson, J., Edwards, J. D., Ackerman, M. L., & Ball, K. (2012). Self-rated Driving and Driving Safety in Older Adults. *Accident Analysis & Prevention*, *48*, 523–527. https://doi.org/10.1038/jid.2014.371

Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E., & Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive Neuroscience*, *1*(3), 165–175. https://doi.org/10.1080/17588921003632529

Sailer, A., Dichgans, J., & Gerloff, C. (2000). The influence of normal aging on the cortical processing of a simple motor task. *Neurology*, *55*(7), 979–985. https://doi.org/10.1212/WNL.55.7.979

Salat, D. H., Kaye, J. A., & Janowsky, J. S. (1999). Prefrontal Gray and White Matter Volumes in Healthy Aging and Alzheimer Disease. *Archives of Neurology*, *56*(3), 338–344. https://doi.org/10.1001/ARCHNEUR.56.3.338

Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological Psychology*, *54*(1–3), 35–54. https://doi.org/10.1016/S0301-0511(00)00052-1

Scheffers, M. K., & Coles, M. G. H. (2000). Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 141–151. https://doi.org/10.1037/0096-1523.26.1.141

Schmidt, J. R. (2019). Evidence against conflict monitoring and adaptation: An updated review. *Psychonomic Bulletin & Review*, *26*(3), 753–771.

Schreiber, M., Pietschmann, M., Kathmann, N., & Endrass, T. (2011). ERP correlates of performance monitoring in elderly. *Brain and Cognition*, *76*(1), 131–139. https://doi.org/10.1016/j.bandc.2011.02.003

Shadlen, M. N., & Kiani, R. (2013). Decision Making as a Window on Cognition. *Neuron, 80*(3), 791–806. https://doi.org/10.1016/J.NEURON.2013.10.047

Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences*, *18*(4), 186–193. https://doi.org/10.1016/j.tics.2014.01.006

Shekhar, M., & Rahnev, D. (2020). The Nature of Metacognitive Inefficiency in Perceptual Decision Making. *Psychological Review*, 1–83. https://doi.org/10.1037/rev0000249

Shimizu, H. (2002). Measuring keyboard response delays by comparing keyboard and joystick inputs. *Behavior Research Methods, Instruments, and Computers*, *34*(2), 250–256. https://doi.org/10.3758/BF03195452

Siedlecka, M., Koculak, M., & Paulewicz, B. (2021). Confidence in action: Differences between perceived accuracy of decision and motor response. *Psychonomic Bulletin and Review*, *28*, 1698–1706. https://doi.org/10.3758/s13423-021-01913-0

Siedlecka, M., Paulewicz, B., & Koculak, M. (2020). Task-related motor response inflates confidence. *BioRxiv*. https://doi.org/10.1101/2020.03.26.010306

Siedlecka, M., Paulewicz, B., & Wierzchoń, M. (2016). But I Was So Sure! Metacognitive Judgments Are Less Accurate Given Prospectively than Retrospectively. *Frontiers in Psychology*, *7*, 1–8. https://doi.org/10.3389/fpsyg.2016.00218

Sim, J., Brown, F., O’Connell, R., & Hester, R. (2020). Impaired error awareness in healthy older adults: an age group comparison study. *Neurobiology of Aging*, *96*, 58–67. https://doi.org/10.1016/j.neurobiolaging.2020.08.001

Siswandari, Y., Bode, S., & Stahl, J. (2019). Performance monitoring beyond choice tasks: The time course of force execution monitoring investigated by event-related potentials and multivariate pattern analysis. *NeuroImage*, *197*, 544–556. https://doi.org/10.1016/j.neuroimage.2019.05.006

Sosnoff, J. J., & Newell, K. M. (2006). *Aging, Visual Intermittency, and Variability in Isometric Force Output*. https://academic.oup.com/psychsocgerontology/article/61/2/P117/558376

Stahl, J., Acharki, M., Kresimon, M., Voller, F., & Gibbons, H. (2015). Perfect error processing: Perfectionism-related variations in action monitoring and error processing mechanisms. *International Journal of Psychophysiology*, *97*(2), 153–162. https://doi.org/10.1016/j.ijpsycho.2015.06.002

Stahl, J., Mattes, A., Hundrieser, M., Kummer, K., Mück, M., Niessen, E., Porth, E., Siswandari, Y., Wolters, P., & Dummel, S. (2020). Neural correlates of error detection during complex response selection: Introduction of a novel eight-alternative response task. *Biological Psychology*, *156*, 107969. https://doi.org/10.1016/j.biopsycho.2020.107969

Stahl, J., & Rammsayer, T. H. (2005). Accessory stimulation in the time course of visuomotor information processing: Stimulus intensity effects on reaction time and response force. *Acta Psychologica*, *120*(1), 1–18. https://doi.org/10.1016/j.actpsy.2005.02.003

Steele, V. R., Anderson, N. E., Claus, E. D., Bernat, E. M., Rao, V., Assaf, M., Pearlson, G. D., Calhoun, V. D., & Kiehl, K. A. (2016). Neuroimaging measures of error-processing: Extracting reliable signals from event-related potentials and functional magnetic resonance imaging. *NeuroImage*, *132*, 247–260. https://doi.org/10.1016/j.neuroimage.2016.02.046

Steinhauser, M., & Yeung, N. (2010). Decision processes in human performance monitoring. *Journal of Neuroscience*, *30*(46), 15643–15653. https://doi.org/10.1523/JNEUROSCI.1899-10.2010

Stone, C., Mattingley, J. B., & Rangelov, D. (2022). On second thoughts: changes of mind in decision-making. *Trends in Cognitive Sciences*, *26*(5), 419–431. https://doi.org/10.1016/J.TICS.2022.02.004

Susser, J. A., & Mulligan, N. W. (2015). The effect of motoric fluency on metamemory. *Psychonomic Bulletin and Review*, *22*(4), 1014–1019. https://doi.org/10.3758/S13423-014-0768-1/FIGURES/1

Tauber, S. K., & Dunlosky, J. (2012). Can older adults accurately judge their learning of emotional information? *Psychology and Aging*, *27*(4), 924–933.

https://doi.org/10.1037/A0028447

Thurm, F., Antonenko, D., Schlee, W., Kolassa, S., Elbert, T., & Kolassa, I. T. (2013). Effects of Aging and Mild Cognitive Impairment on Electrophysiological Correlates of Performance Monitoring. *Journal of Alzheimer's Disease*, *35*(3), 575–587. https://doi.org/10.3233/JAD-121348

Thurm, F., Li, S. C., & Hämmerer, D. (2020). Maturation- and aging-related differences in electrophysiological correlates of error detection and error awareness. *Neuropsychologia*, *143*. https://doi.org/10.1016/j.neuropsychologia.2020.107476

Turner, W., Angdias, R., Feuerriegel, D., Chong, T. T., Hester, R., & Bode, S. (2021). Perceptual decision confidence is sensitive to forgone physical effort expenditure. *Cognition*, *207,* 104525. https://doi.org/https://doi.org/10.1016/j.cognition.2020.104525

Turner, W., Feuerriegel, D., Andrejević, M., Hester, R., & Bode, S. (2021). Perceptual change-of-mind decisions are sensitive to absolute evidence magnitude. *Cognitive Psychology*, *124,* 101358. https://doi.org/10.1016/j.cogpsych.2020.101358

Turner, W., Feuerriegel, D., Hester, R., & Bode, S. (2022). An initial "snapshot" of sensory information biases the likelihood and speed of subsequent changes of mind. *PLoS Computational Biology*, *18*(1). https://doi.org/10.1371/journal.pcbi.1009738

Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience*, *42*(1), 1636–1643. https://doi.org/10.1111/EJN.12936

Ullsperger, M., Fischer, A. G., Nigbur, R., & Endrass, T. (2014). Neural mechanisms and temporal dynamics of performance monitoring. *Trends in Cognitive Sciences*, *18*(5), 259–267. https://doi.org/10.1016/j.tics.2014.02.009

Ullsperger, M., Harsay, H. A., Wessel, J. R., & Ridderinkhof, K. R. (2010). Conscious perception of errors and its relation to the anterior insula. *Brain Structure and Function*, *214*(5), 629–643. https://doi.org/10.1007/S00429-010-0261-1

Ulrich, R., Rinkenauer, G., & Miller, J. (1998). Effects of Stimulus Duration and Intensity on Simple Reaction Time and Response Force. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 915–928.

Ulrich, R., & Wing, A. M. (1991). A Recruitment Theory of Force-Time Relations in thte Production of Brief Force Pulses: The Parallel Force Unit Model. *Psychological Review*, *98*(2), 268–294. https://doi.org/https://doi.org/10.1037/0033-295X.98.2.268

Van Den Berg, R., Anandalingam, K., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). A common mechanism underlies changes of mind about decisions and confidence. *ELife*, *5*, 1–21. https://doi.org/10.7554/eLife.12192

Van Der Lubbe, R. H. J., Verleger, R., der Lubbe, R. H. J., & Verleger, R. (2002). Aging and the Simon task. *Psychophysiology*, *39*(1), 100–110. https://doi.org/https://doi.org/10.1111/1469-8986.3910100

van Gaal, S., de Lange, F. P., & Cohen, M. X. (2012). The role of consciousness in cognitive control and decision making. *Frontiers in Human Neuroscience*, *0*, 121. https://doi.org/10.3389/FNHUM.2012.00121/BIBTEX

Verkuilen, J., & Smithson, M. (2012). Mixed and mixture regression models for continuous bounded responses using the beta distribution. *Journal of Educational and Behavioral Statistics*, *37*(1), 82–113. https://doi.org/10.3102/1076998610396895

Vickers, D. (1979). *Decision Processes in Visual Perception*. Academic Press.

Vickers, D., & Packer, J. (1982). Effects of alternating set for speed or accuracy on response time, accuracy and confidence in a unidimensional discrimination task. *Acta Psychologica*, *50*(2), 179–197. https://doi.org/10.1016/0001-6918(82)90006-3

Vidal, F., Burle, B., Bonnet, M., Grapperon, J., & Hasbroucq, T. (2003). Error negativity on correct trials: a reexamination of available data. *Biological Psychology*, *64*, 265–282. https://doi.org/10.1016/S0301-0511(03)00097-8

Vissers, M. E., Ridderinkhof, K. R., Cohen, M. X., & Slagter, H. A. (2018). Oscillatory mechanisms of response conflict elicited by color and motion direction: An individual differences approach. *Journal of Cognitive Neuroscience*, *30*(4), 468–481.

Vukman, K. B. (2005). Developmental Differences in Metacognition and their Connections with Cognitive Development in Adulthood. *Journal of Adult Development*, *12*(4), 211–221. https://doi.org/10.1007/S10804-005-7089-6

Weil, L. G., Fleming, S. M., Dumontheil, I., Kilford, E. J., Weil, R. S., Rees, G., Dolan, R. J., & Blakemore, S. J. (2013). The development of metacognitive ability in adolescence. *Consciousness and Cognition*, *22*(1), 264–271. https://doi.org/10.1016/J.CONCOG.2013.01.004

Wessel, J. R. (2012). Error awareness and the error-related negativity: evaluating the first decade of evidence. *Frontiers in Human Neuroscience*, *6*, 1–16. https://doi.org/10.3389/fnhum.2012.00088

Wessel, J. R. (2018). An adaptive orienting theory of error processing. *Psychophysiology*, *55*(3), 1–21. https://doi.org/10.1111/psyp.13041

Wessel, J. R., Danielmeier, C., & Ullsperger, M. (2011). Error awareness revisited: Accumulation of multimodal evidence from central and autonomic nervous systems. *Journal of Cognitive Neuroscience*, *23*(10), 3021–3036. https://doi.org/10.1162/jocn.2011.21635

Wessel, J. R., Dolan, K. A., & Hollingworth, A. (2018). A blunted phasic autonomic response to errors indexes age-related deficits in error awareness. *Neurobiology of Aging*, *71*, 13–20. https://doi.org/10.1016/j.neurobiolaging.2018.06.019

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin*, *120*(2), 272–292. https://doi.org/10.1037/0033-2909.120.2.272

Wilson, R. S., Boyle, P. A., Yu, L., Barnes, L. L., Sytsma, J., Buchman, A. S., Bennett, D. A., & Schneider, J. A. (2015). Temporal course and pathologic basis of unawareness of memory loss in dementia. *Neurology*, *85*(11), 984–991. https://doi.org/10.1212/WNL.0000000000001935

Wokke, M. E., Achoui, D., & Cleeremans, A. (2020). Action information contributes to metacognitive decision-making. *Scientific Reports*, *10*(1), 1–15. https://doi.org/10.1038/s41598-020-60382-y

World Bank Open Data. (n.d.). https://data.worldbank.org/

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychological Review*, *111*(4), 931–959. https://doi.org/10.1037/0033-295X.111.4.931

Yeung, N., & Cohen, J. D. (2006). The impact of cognitive deficits on conflict monitoring predictable dissociations between the error-related negativity and N2. *Psychological Science*, *17*(2), 164–171. https://doi.org/10.1111/j.1467-9280.2006.01680.x

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1310–1321. https://doi.org/10.1098/rstb.2011.0416

Yordanova, J., Kolev, V., Hohnsbein, J., & Falkenstein, M. (2004). Sensorimotor slowing with ageing is mediated by a functional dysregulation of motor-generation processes: evidence from high-resolution event-related potentials. *Brain*, *127*(2), 351–362. https://doi.org/10.1093/brain/awh042

Zakrzewski, A. C., Sanders, E. C., & Berry, J. M. (2021). Evidence for Age-Equivalent and Task-Dissociative Metacognition in the Memory Domain. *Frontiers in Psychology*, *12*, 57. https://doi.org/10.3389/FPSYG.2021.630143/BIBTEX